THOMSON

1ª PARTE

INSTRUMENTACIÓN ELECTRÓNICA

INCLUYE CD

Miguel A. Pérez García
Juan C. Álvarez Antón
Juan C. Campo Rodríguez
Fco. Javier Ferrero Martín
Gustavo J. Grillo Ortega

INSTRUMENTACION ELECTRÓNICA

Miguel A. Pérez García
Juan C. Álvarez Antón
Juan C. Campo Rodríguez
Fco. Javier Ferrero Martín
Gustavo J. Grillo Ortega

Escuela Politécnica Superior de Ingeniería de Gijón
Universidad de Oviedo

THOMSON™

Australia • Canadá • México • Singapur • España • Reino Unido • Estados Unidos

Contenido

Capítulo 1

Introducción a la instrumentación electrónica 1

1.1 Introducción 1
1.1.1. Variables y señales 2
1.1.2. La instrumentación electrónica en el control de procesos 5
1.2 Sistemas de medida 9
1.2.1. Funciones de un sistema de medida 10
1.2.2. Sistemas de medida multicanal 11
1.2.3. Arquitecturas de los sistemas de instrumentación 13
1.3 Características estáticas 15
1.3.1. Curva de calibración (*Static Transfer Function*) 15
1.3.2. Errores: exactitud, veracidad y precisión 20
1.3.3. Propagación de errores 21
1.3.4. Calibración 23
1.4 Características dinámicas 25
1.4.1. Función de transferencia 25
1.4.2. Caracterización de la función de transferencia 28
1.4.3. Evaluación de la respuesta dinámica 32

Capítulo 2

Amplificación 41

2.1 Conceptos generales 41
2.1.1. Amplificar y modificar los valores de una señal 46
2.1.2. Ruido generado 46
2.1.3. Filtrado 47
2.2 El amplificador operacional 48
2.2.1. El amplificador operacional ideal 48
2.2.2. El amplificador operacional realimentado en tensión 51
2.2.3. Parámetros reales de los operacionales 52
2.2.3.1. Características estáticas de los amplificadores operacionales 55
2.2.3.2. Características dinámicas de los amplificadores operacionales 76
2.2.3.3. Ruido interno 86

Capítulo 3

Circuitos amplificadores de uso en instrumentación 99

3.1 Amplificadores de instrumentación 99
3.1.1. Particularidades del amplificador de instrumentación 100
3.1.2. Parámetros que caracterizan a un amplificador de instrumentación 110

3.2 Amplificadores aislados o de aislamiento ... 115
3.2.1. Tipos de amplificadores aislados ... 116
3.2.2. Parámetros característicos de los amplificadores de aislamiento ... 120
3.2.3. Aplicaciones de los amplificadores aislados ... 123
3.3 Amplificadores de transimpedancia y transconductancia ... 123
3.3.1. Amplificador de transimpedancia ... 124
3.3.2. Amplificador de transconductancia ... 126
3.4 Otros tipos de amplificadores ... 128
3.4.1. Amplificadores troceadores ... 128
3.4.2. Amplificadores logarítmicos ... 130
3.5 Algunos criterios de selección de amplificadores operacionales para instrumentación 134
3.6 Otras consideraciones en el diseño de circuitos con amplificadores operacionales ... 136
3.6.1. Circuitos basados en operacionales con resistencias reales ... 137
3.6.2. Circuitos integradores ... 139
3.6.3. Circuitos derivadores ... 141

Capítulo 4
Filtros analógicos ... 145

4.1 Introducción ... 145
4.2 Clasificación de filtros ... 146
4.3 Especificaciones de un filtro ... 148
4.3.1. Especificaciones frecuenciales ... 148
4.3.2. Especificaciones temporales ... 150
4.4 Aproximaciones matemáticas ... 151
4.5 Filtros pasivos RC ... 156
4.6 Filtros activos ... 161
4.6.1. Estructuras para la realización de filtros activos ... 161
4.6.2. Diseño de filtros paso bajo ... 161
4.6.3. Diseño de filtros paso alto ... 169
4.6.4. Diseño de filtros paso banda y rechazo de banda ... 170
4.7 Filtros activos de variables de estado ... 173
4.8 Filtros de capacidad conmutada ... 175
4.9 Selección de componentes ... 179
4.10 Programas de ayuda al diseño de filtros ... 182
4.11 Comparación de tecnologías de filtros ... 183
4.12 Aplicaciones de los filtros ... 184

Capítulo 5
Sensores potenciométricos ... 191

5.1 Características generales ... 191
5.2 Parámetros característicos de los potenciómetros ... 193
5.3 Tipos de potenciómetros ... 194
5.4 Acondicionamiento de señal en potenciómetros ... 198
5.5 Errores debidos al cableado ... 203

Capítulo 6
Sensores de temperatura de resistencia metálica ... 207

6.1 Características generales ... 207
6.1.1. Curvas de calibración ... 208
6.1.2. Modelo matemático de la RTD ... 210

6.2 Tipos de RTD. Comparación general 210
6.3 El autocalentamiento 212
6.4 Aplicaciones 213
6.5 Acondicionamiento de la señal 213
6.5.1. Puente de Wheatstone alimentado con tensión 213
6.5.1.1. Linealidad 214
6.5.1.2. Sensibilidad 215
6.5.1.3. Conexión remota del puente 217
6.5.2. Puente de Wheatstone alimentado por corriente 220
6.5.2.1. Linealidad 220
6.5.2.2. Sensibilidad 221
6.5.3. Otros circuitos acondicionadores para RTDs 223

Capítulo 7
Galgas extensométricas 227

7.1 Principio de funcionamiento 227
7.2 Tipos de galgas extensométricas 230
7.3 Utilización de las galgas extensométricas 232
7.4 Circuitos de medida 234
7.5 Utilización de los circuitos de medida 239
7.6 Aplicaciones 242
7.6.1. Medida de estados de deformación 242
7.6.2. Medidas de otras variables con galgas extensométricas 245

Capítulo 8
Termistores y fotorresistencias 249

8.1 Introducción 249
8.2 Termistores: NTCs 249
8.2.1. Característica R-T de una NTC 251
8.2.2. Aproximaciones de la característica resistencia-temperatura 251
8.2.3. La NTC como elemento de circuito 252
8.2.4. La NTC como sensor de temperatura 254
8.2.5. Otras aplicaciones de las NTCs 258
8.3 Termistores de coeficiente de temperatura positivo (PTCs) 262
8.3.1. Principio físico de funcionamiento 262
8.3.2. Características generales 262
8.3.3. Aplicaciones 266
8.4 Fotorresistencias (LDRs) 268
8.4.1. Principio de funcionamiento 268
8.4.2. Modelo de la LDR 270
8.4.3. Aplicaciones de la LDR 271

Capítulo 9
Otros sensores resistivos 277

9.1 Sensores de gases de óxidos metálicos semiconductores 277
9.2 Magnetorresistencias 281

Capítulo 10
Sensores capacitivos 285

10.1 Introducción 285
10.2 Variación de la capacidad en un condensador de placas paralelas 287
10.2.1. Condensador simple 287
10.2.2. Condensador diferencial 287
10.3 Circuitos de medida 289
10.3.1. Señal de excitación 290
10.3.2. Amplificador de alterna 290
10.3.3. Puentes de alterna 291
10.3.4. Oscilador de frecuencia variable 292
10.3.5. Demodulador síncrono 292
10.3.6. Circuitos de capacidad conmutada 293
10.4 Detectores de proximidad capacitivos 295
10.5 Sensores capacitivos en silicio 297

Capítulo 11
Sensores inductivos 305

11.1 Introducción 305
11.2 Sensores inductivos básicos 307
11.2.1. Sistemas con sensores inductivos 307
11.2.2. Circuitos de medida 309
11.2.3. Sistemas de medida con sensores inductivos 311
11.3 El transformador diferencial lineal (LVDT) 311
11.3.1. Descripción de funcionamiento 311
11.3.2. Circuitos de medida 312
11.3.3. Consideraciones prácticas en el diseño de circuitos con LVDT 314
11.4 Otros sensores inductivos 318

Capítulo 12
Sensores electromagnéticos 325

12.1 Introducción 325
12.2 Sensores electromagnéticos lineales 326
12.3 Sensores electromagnéticos rotativos 326
12.3.1. Sensores electromagnéticos rotativos de velocidad angular 327
12.3.2. Sensores electromagnéticos rotativos de posición angular 329

Capítulo 13
Termopares 333

13.1 Principio de funcionamiento 333
13.2 Tipos de termopares 335
13.3 Curvas de calibración 337
13.4 Efectos de las uniones parásitas 341
13.5 Acondicionamiento de señal 341

Capítulo 14
Sensores piezoeléctricos 347

14.1 Introducción 347
14.2 Comportamiento de los materiales piezoeléctricos 350
14.3 Los dispositivos piezoeléctricos como sensores 353
14.3.1. Medida de fuerza, presión y aceleración 353
14.3.2. Problemática de la utilización de sensores piezoeléctricos 355
14.4 Sensores de ultrasonidos 364
14.4.1. Los ultrasonidos 365
14.4.2. Técnicas de impulso-eco 365
14.4.3. Aplicaciones de las técnicas de impulso-eco 370

Capítulo 15
Sensores piroeléctricos 377

15.1 Introducción 377
15.2 Principio de funcionamiento 377
15.3 Circuito equivalente de un sensor piroeléctrico 379
15.4 Acondicionamiento de señal 381
15.5 Aplicaciones 383

Capítulo 16
Sensores optoelectrónicos generadores de señal 387

16.1 Fotodiodos y fototransistores 387
16.1.1. Principio de funcionamiento de los fotodiodos 387
16.1.2. Tipos de fotodiodos 390
16.1.3. Modelo eléctrico de un fotodiodo 392
16.1.4. Características de los fotodiodos 393
16.1.5. Acondicionamiento de fotodiodos 397
16.1.5.1. Generalidades 397
16.1.5.2. Amplificador corriente-tensión (amplificador de transimpedancia) 399
16.1.5.3. Problemática del amplificador corriente-tensión 403
16.1.6. Fototransistores 405
16.1.7. Aplicaciones de fotodiodos y fototransistores 406
16.1.7.1. Detectores de proximidad fotoeléctricos 407
16.1.7.2. Codificadores ópticos (*optical encoders*) 408
16.1.7.3. Sensores de color 411
16.1.7.4. Detectores de humo y turbidímetros 411
16.1.7.5. Espectrofotometría de absorción 412
16.2 Dispositivos de acoplamiento de carga (CCDs) 413
16.2.1. Principio de funcionamiento 413
16.2.1.1. Conversión de la luz a carga eléctrica 414
16.2.1.2. Transferencia de las cargas 415
16.2.1.3. Conversión de la carga a tensión (etapa de salida) 417
16.2.2. Arquitecturas 419
16.2.3. Características 421
16.2.4. CCDs para detección de imágenes en color 422
16.3 Fotomultiplicadores 423
16.3.1. Principio de funcionamiento 423
16.3.2. Tipos de fotomultiplicadores 423
16.3.3. Características 424

16.3.4. Alimentación 425
16.3.5. Acondicionamiento de señal 426
16.3.6. Aplicaciones 427

Capítulo 17
Sensores de efecto Hall 431

17.1 El efecto Hall 431
17.2 Sensores Hall de salida lineal 433
17.3 Sensores Hall de salida digital 436
17.4 Modos de operación 438
17.5 Dispositivos de medida basados en el efecto Hall 440
17.6 Ejemplos de aplicaciones 442

Capítulo 18
Otros tipos de sensores 451

18.1 Sensores de fibra óptica 451
18.1.1. Fundamentos de la fibra óptica 451
18.1.2. Clasificación de los sensores de fibra óptica 456
18.1.3. Sensores de fibra óptica basados en modulación de amplitud 457
18.1.4. Sensores de fibra óptica basados en modulación de fase 461
18.2 Biosensores 464
18.2.1. Introducción 465
18.2.2. Tipos de biosensores 466

Capítulo 19
Criterios para la selección de sensores 473

19.1 Medida de la temperatura 473
19.1.1. Introducción 473
19.1.2. RTD vs termopar 474
19.1.3. NTC vs sensores de silicio 477
19.1.4. Sensores de radiación 478
19.2 Medida de las principales variables mecánicas 481
19.2.1. Medida de presencia y proximidad 482
19.2.2. Medida de la posición 486
19.2.3. Medida de la velocidad 493
19.2.4. Medida de la aceleración y de la vibración 496
19.2.5. Medida de fuerza y de peso 498
19.2.6. Medida de la presión 501
19.2.7. Medida del flujo 507
19.2.8. Medida del nivel 513
19.3 Medida de magnitudes eléctricas 519
19.3.1. Resistencia *shunt* 519
19.3.2. Transformador de intensidad 521
19.3.3. Sensores de efecto Hall 523
19.3.4. Sensores magnetorresistivos 526
19.3.5. Sensores de fibra óptica 528
19.4 Medida de magnitudes ópticas 530
19.4.1. Sistemas de medida de radiación óptica 530
19.4.2. Sensores ópticos 532

Capítulo 20
Introducción a la transmisión de señal 537

20.1 La transmisión 537
20.1.1. Medios de transmisión 538
20.1.1.1. Medios de transmisión guiados 539
20.1.1.2. Medios no guiados 546
20.1.2. Codificación de la información 549
20.2 Telemedida 550
20.3 Bucles de tensión y bucles de corriente 551
20.3.1. Bucles de tensión 551
20.3.2. Bucles de corriente 555
20.3.3. Conversión V/I 559
20.3.4. Conversión I/V 562
20.3.5. Consideraciones prácticas sobre los bucles de corriente 563

Capítulo 21
Modulación y demodulación 567

21.1 Modulación con portadora analógica y moduladora analógica 569
21.1.1. Modulación AM 569
21.1.2. Modulación FM 573
21.1.3. Modulación PM 576
21.2 Modulación con portadora analógica y moduladora digital 577
21.2.1. Modulación ASK 577
21.2.2. Modulación FSK 580
21.2.3. Modulación PSK 583
21.3 Modulaciones con portadora digital y moduladora analógica 586

Capítulo 22
Variables muestreadas 591

22.1 Introducción 591
22.2 Variables analógicas, digitales y muestreadas 591
22.3 Teorema del muestreo 592
22.3.1. El problema del solapamiento o *aliasing* 596
22.3.2. Selección de la frecuencia de muestreo 597
22.3.3. Limitaciones 598
22.4 Variables submuestreadas 601
22.5 Reconstrucción de señales a partir de muestras 601
22.6 Muestreo y retención 603
22.6.1. El modo muestreo 603
22.6.2. El modo retención 604
22.6.3. La transición del modo muestreo al modo retención 605
22.6.4. Circuitos prácticos de muestreo y retención 607
22.7 Multiplexación de señales analógicas 609
22.7.1. La multiplexación y la frecuencia de muestreo 611

Capítulo 23
Conversión entre variables analógicas y digitales 617

23.1 Cuantificación 617
23.2 Codificación 621
23.2.1. Códigos binarios unipolares 621
23.2.2. Códigos binarios bipolares 622

23.3 Conversión Digital-Analógica 624
23.3.1. Características de la conversión D/A 625
23.3.1.1. Características estáticas 625
23.3.1.2. Características dinámicas 626
23.3.1.3. Errores en los convertidores D/A 628
23.3.2. Tipos de convertidores D/A 631
23.3.2.1. Convertidores de elementos ponderados en binario 631
23.3.2.2. Convertidores de código de termómetro 634
23.3.2.3. Convertidores de redes escalera (*ladder*) 637
23.3.2.4. Convertidores de tensiones o corrientes segmentadas 639
23.3.2.5. Convertidores Sigma-Delta 641
23.3.3. Estructuras de la entrada en los circuitos integrados. Circuitos múltiples 642
23.3.3.1. Estructuras de la entrada 642
23.3.3.2. Convertidores múltiples 643
23.3.4. Comparación y criterios de selección 643
23.4 Conversión Analógica-Digital 644
23.4.1. Características de la conversión A/D 645
23.4.1.1. Características estáticas 645
23.4.1.2. Características dinámicas 648
23.4.1.3. Errores en los convertidores A/D 649
23.4.2. Tipos de convertidores 654
23.4.2.1. Convertidores directos 654
23.4.2.2. Convertidores indirectos 659
23.4.3. Comparación y criterios de selección 669

Capítulo 24
Procesadores digitales de señal 673

24.1 Microprocesadores, microcontroladores y DSP 673
24.2 Procesamiento de la información 692
24.3 Comunicación hombre-máquina 693
24.3.1. Presentación de la información 693
24.4 Sensores inteligentes 700

Capítulo 25
Sistemas de adquisición de datos 707

25.1 Introducción a los sistemas de adquisición de datos (SAD) 707
25.2 Configuraciones de un sistema de adquisición de datos 708
25.2.1. Tarjetas de adquisición de datos 709
25.3 Buses de instrumentos 716
25.3.1. El bus GPIB 717
25.3.2. Bus VXI 721

Capítulo 26
Buses de campo 725

26.1 Introducción: conceptos previos 725
26.2 Características de las redes de control 729
26.3 Buses de campo 732

Capítulo 27
Software de instrumentación 741

27.1 Instrumentos digitales 741
27.2 Software de instrumentación 741

Capítulo 28
Interferencias electromagnéticas 747

28.1 Introducción 747
28.1.1. Definiciones 747
28.1.2. Desde la fuente a la víctima 749
28.1.3. Normativa 752
28.2 Fuentes de interferencias 752
28.2.1. Sistemas y componentes sin transitorios 753
28.2.2. Sistemas y componentes con transitorios 755
28.2.3. Sistemas generadores de arcos 756
28.2.4. Otras fuentes de interferencias 758
28.3 Acoplamiento de fuentes de interferencias 759
28.3.1. Acoplamientos conductivos 759
28.3.2. Acoplamientos no conductivos 763
28.4 Minimización de los efectos de las interferencias 770

Capítulo 29
Cableado y apantallado 773

29.1 Minimización de interferencias conductivas 773
29.1.1. Conexionado de las masas y alimentaciones 773
29.1.2. Problemática de la conexión a tierra 775
29.1.2.1. Bucles de tierra 775
29.1.2.2. Punto de conexión a tierra 781
29.2 Minimización de interferencias debidas a acoplamientos inductivos 783
29.3 Minimización de interferencias debidas a acoplamientos capacitivos 784
29.3.1. Guardas activas 786
29.4 Minimización de interferencias radiadas. Pantallas 789
29.4.1. Pantallas para campos cercanos 791

Capítulo 30
Las tarjetas de circuito impreso 797

30.1 Aspectos generales en el diseño de circuitos electrónicos y normativa 797
30.2 Tarjetas de circuito impreso. Tecnologías 801
30.3 Consideraciones de diseño de los circuitos impresos para sistemas de instrumentación 803
30.3.1. Las alimentaciones y la masa 805
30.3.2. Las pistas de señal 810
30.3.2.1. Trazado de pistas para señales referidas a masa 810
30.3.2.2. Pistas para señales diferenciales 818
30.4 ¿Qué debemos tener en cuenta? 820

Capítulo 31

Seguridad en los sistemas de instrumentación 823

31.1 La puesta a tierra 823
31.2 Seguridad en atmósferas explosivas 826
31.2.1. Métodos de protección 828
31.2.2. Seguridad intrínseca 829
31.2.2.1. Riesgo de inflamación debido a chispas 829
31.2.2.2. Riesgo de inflamación debido a temperaturas elevadas 833
31.2.2.3. Interfaz entre circuitos de seguridad intrínseca y circuitos sin seguridad intrínseca ... 833
31.2.2.4. La puesta a tierra 835
31.2.2.5. Certificación 836

Bibliografía y referencias 839

Índice alfabético 853

Introducción

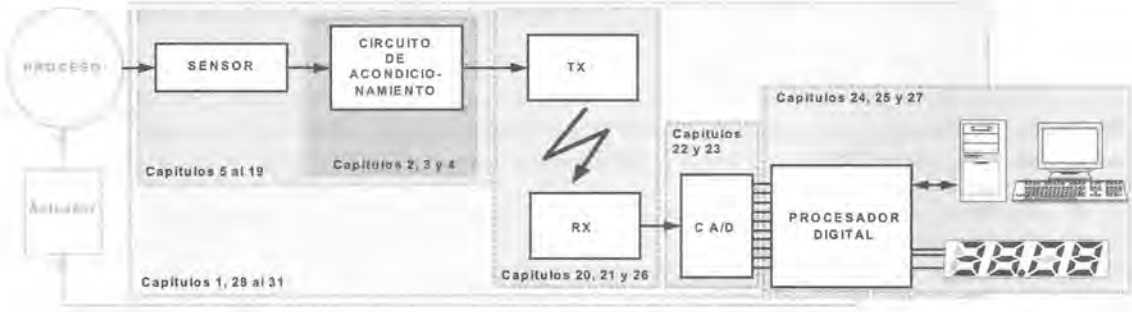
Sobre la estructura del libro

El presente libro de instrumentación electrónica tiene como objetivo establecer conceptos generales sobre los sistemas de instrumentación electrónica sea cual sea su ámbito de aplicación. Sin embargo, dado el carácter globalizador de los sistemas de instrumentación en el campo del control de procesos, hemos empleado este caso como marco de todo el libro.

Tratando de huir de un tono enciclopédico se han desarrollado diversos temas que configuran en su conjunto un sistema de instrumentación genérico. Para no alargar la obra, en el tintero se han quedado algunos aspectos y el lector echará en falta que no se trate de determinado sensor, aplicación o circuito de tratamiento; sin embargo, esperamos que el poso de conocimiento que quede tras la lectura del libro sea suficiente para abordar otros temas no tratados o aquellos sistemas que sencillamente no existen a fecha de hoy.

Nuestro objetivo ha sido estimular la crítica y el interés por la indagación del lector en un tema de índole horizontal, en progreso constante como es la Instrumentación Electrónica y de tan amplias miras que afecta a la mayoría de las facetas de nuestra vida diaria, desde el hogar hasta los lugares de trabajo.

Para conseguirlo, el contenido de la obra se estructura en 31 capítulos y un CD con información adicional, pero no menos relevante. Como al principio se indicaba, usando como marco el control de procesos se ha presentado un sistema de instrumentación electrónica según se indica en la figura.



En el capítulo 1 se dan las ideas básicas de carácter general de la Instrumentación Electrónica y que se usarán después a lo largo de los demás capítulos; el lector debe prestar un especial interés a este capítulo. En los capítulos 2 al 4 se tratan los sistemas de tratamiento de señal de tipo analógico que usarán los circuitos de acondicionamiento de los diversos sensores tratados en los capítulos 5 al 18.

Hemos optado por una clasificación horizontal de los sensores, por su principio físico de funcionamiento y no por la variable a medir porque creemos que esto facilita tanto la comprensión de su modo de trabajo como el estudio de los necesarios circuitos de acondicionamiento cuyas bases ya están establecidas. Sin embargo, creemos que es necesario también hacer una recapitulación por variable a medir: en el capítulo 19 se realiza una clasificación de tipo vertical, estableciendo los criterios para seleccionar sensores para una medida concreta, sea cual sea su principio físico de trabajo. El lector podrá usar este capítulo a modo de resumen de los anteriores, como reforzamiento de los conceptos ya manejados y como pequeña guía de diseño.

La necesidad de transmisión de la información procedente de los sensores hasta los puntos de lectura o control se trata en los capítulos 20, 21 y 26 de cuyo conjunto se puede extraer el concepto de telemedida como uno de los aspectos que caracteriza la Instrumentación Electrónica hoy en día.

Ya sea por necesidades de control ya sea por la necesidad de ajustes, modificación de la información o por su distribución, la mayoría de los datos de un sistema de instrumentación acaban en algún tipo de procesador digital. Para tratar estos aspectos se han dedicado los capítulos 22 y 23 al paso de señales analógicas al dominio digital y los capítulos 24, 25 y 27 al tratamiento digital de los datos, haciendo especial hincapié en los sistemas instrumentales digitales basados en computador.

Para concluir el libro se vuelve al concepto general de sistema de instrumentación para tratar aspectos tan relevantes como las interferencias, la seguridad y el diseño optimizado de los sistemas desde este punto de vista, tanto en el nivel de interconexión de bloques, equipos y subsistemas como en el trazado de tarjetas de circuito impreso.

Cada uno de los capítulos se ha desarrollado teóricamente pero obviando los desarrollos matemáticos, que el lector muy interesado podrá encontrar en la bibliografía que se detalla al final del texto, e incluyendo un buen número de ejercicios y ejemplos que contribuyen a hacer más amena la lectura y permiten comprobar la asimilación de los conceptos más importantes. Aún está disponible una comprobación más, por medio de los problemas propuestos; todos ellos están completamente resueltos en el CD adjunto pero invitamos al lector a tratar de resolverlos por su cuenta, acudiendo a las soluciones para contrastar los resultados obtenidos.

También se suministran en el CD adjunto todas las hojas de características de los dispositivos que se mencionan en el texto del libro y que el lector debe consultar para comprobar que es capaz de asimilar la mayoría de los datos que figuran en ellas y conceder la importancia que tengan en cada caso o aplicación. También el lector podrá buscar en Internet —o en otra fuente como catálogos impresos— nuevas hojas de datos bien sea usando los enlaces que se proporcionan en el CD adjunto o utilizando las herramientas clásicas de búsqueda.

La curiosidad puede hacer interesante conocer cómo son determinados componentes o sistemas y el aspecto que tienen; para satisfacer esta curiosidad se han proporcionado en el CD adjunto unas cuantas fotografías de sensores y sistemas que ayudarán a una mejor identificación y a una cierta familiarización con ellos.

Para concluir, queremos destacar un aspecto que consideramos como muy importante: se han realizado cinco diseños completos de sistemas de instrumentación de diversa índole que el lector podrá consultar en el CD adjunto. No son "problemas" en el sentido clásico que la palabra tiene en el entorno docente, sino que constituyen ejercicios de recapitulación de una buena parte del texto. Invitamos al lector a que los siga con detalle y con una actitud crítica; como ejemplos de diseño que son, el lector debería ser capaz de proporcionar otras alternativas, cuestionar las decisiones tomadas, valorar su ámbito de validez, en fin, aprovechar los conocimientos que haya obtenido para alcanzar el nivel de síntesis en la escala del aprendizaje.

En el CD adjunto hemos incluido unas presentaciones muy sencillas en PowerPoint de todos los capítulos de este texto, incluyendo la mayoría de las figuras que forman parte del libro. Creemos que puede ser de

alguna ayuda bien para poder usarlas como están, bien para modificarlas añadiendo la información que estime oportuna en la exposición de la materia a los alumnos. Si desea incluir alguna o parte de las figuras de esas presentaciones o de las fotografías que se incluyen en el CD en cualquier otro documento recordamos la necesidad de cumplir con las condiciones del copyright.

A pesar del cuidado que hemos tenido en la elaboración del texto y en la resolución de los ejercicios y problemas somos conscientes que se habrán colado algunos gazapos. Estaríamos muy agradecidos nos hagan llegar cuantos comentarios, sugerencias y críticas tengan a bien hacernos. Solo así podremos mejorar el libro.

Finalmente agradecer a Dª Patricia Roldán (Universidad de Oviedo) y a D. Diego Gago (DME S.L.) el esfuerzo y dedicación que han puesto en la elaboración de las figuras que forman parte del libro y que se incluyen en las transparencias.

1

Introducción a la instrumentación electrónica

1.1 Introducción
1.2 Sistemas de medida
1.3 Características estáticas
1.4 Características dinámicas

1.1 Introducción

La medida consiste en la determinación de una magnitud por comparación con un estándar. Aunque algunos de nuestros sentidos puedan parecernos, a priori, extraordinariamente desarrollados, la percepción y medida de magnitudes físicas no puede realizarse con precisión. En primer lugar, no todos percibimos de la misma forma, por lo que se hace muy difícil cuantificar con objetividad. Además, los sentidos son incapaces de estimar ciertas variables físicas. Así, por citar algunos ejemplos, no podemos percibir ondas sonoras de una frecuencia superior a 20 kHz, ni tampoco ondas electromagnéticas que salgan fuera de nuestro reducido espectro visible. Para suplir estas deficiencias se recurre a los instrumentos de medida. Los instrumentos son capaces de cuantificar de forma sistemática y de "ver" lo que el hombre no es capaz de ver. Un microscopio, por ejemplo, es un instrumento óptico que nos permite reconocer objetos extraordinariamente pequeños y una simple cinta métrica, otro instrumento del que nos valemos para estimar distancias.

El ser humano percibe la información del mundo que le rodea a través de sus sentidos y adquiere el conocimiento, sobre todo el científico, cuando es capaz de cuantificar las magnitudes que percibe, es decir, a través de la medida. Según J. Ferrero [1.17]: "La instrumentación comprende todas las técnicas, equipos y metodologías relacionadas con el diseño, la construcción y la aplicación de dispositivos físicos para mejorar, completar y aumentar la eficiencia de los mecanismos de percepción del ser humano".

Entre los instrumentos, gozan de especial interés aquellos que utilizan técnicas electrónicas para realizar la medida. Por ello, una de las tecnologías de instrumentación más avanzadas es la denominada **Instrumentación Electrónica** que es la técnica que se ocupa de la medición de cualquier tipo de magnitud física, de la conversión de la misma a magnitudes eléctricas y de su tratamiento para proporcionar la información adecuada a un sistema de control, a un operador humano o a ambos. La instrumentación electrónica tiene su campo de aplicación en numerosas actividades relacionadas con la ciencia y la tecnología donde la electrónica se incorporó de forma masiva. En el laboratorio y en la industria se realiza un uso intensivo de la instrumentación electrónica porque en estos entornos, como en otros muchos, es necesario medir para conocer.

1.1.1 VARIABLES Y SEÑALES

La información de las variables que se pretenda capturar se almacena en algún tipo de variable eléctrica, generalmente tensión. Esa variable eléctrica es lo que se denomina **señal**. La naturaleza de las variables y de las señales que las contienen puede ser igual o distinta: en el primer caso, variable y señal coinciden (o son proporcionales) mientras que, en el segundo, la variable es almacenada "dentro" de alguno de los parámetros de la señal. En cualquiera de los casos, variables y señales pueden clasificarse siguiendo varios criterios, pero aquí nos fijaremos en aquellos que resultan interesantes desde el punto de vista de la instrumentación electrónica. De esta forma, tenemos:

a) **Variables analógicas** cuando los datos constituyen matemáticamente un conjunto denso[1], es decir, que puede tener cualquier valor dentro de un intervalo determinado según se indica en la figura 1.1a, lo que implica que el número total de valores es infinito.

b) **Variables digitales** cuando los datos constituyen un conjunto finito de valores; un caso particular de este tipo de variables es el sistema binario que permite dos valores diferentes, denotados normalmente por "0" y "1" (figura 1.1b).

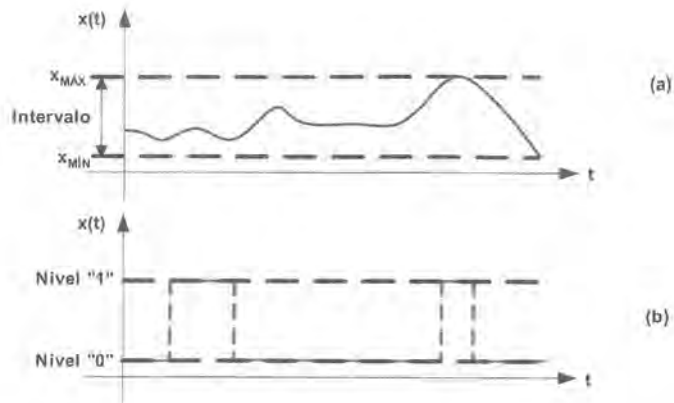


Figura 1.1. Tipos de variables. Representación de: (a) una variable analógica y (b) una variable digital binaria.

Las señales, igual que las variables, pueden clasificarse según el mismo criterio, obteniendo señales analógicas y señales digitales. Tengamos en cuenta que una señal digital o una señal analógica puede contener

[1] Un conjunto denso de datos es aquel que tiene la particularidad de que entre dos valores cualesquiera de él, se puede incluir otro valor del mismo conjunto.

variables (informaciones) tanto digitales como analógicas. Esto se observa mejor en las figuras 1.2a y 1.2b donde se muestra una variable analógica o en la figura 1.2d, en que aparece una variable digital: ambas pueden incluir diversas variables cuya naturaleza puede ser analógica o digital.

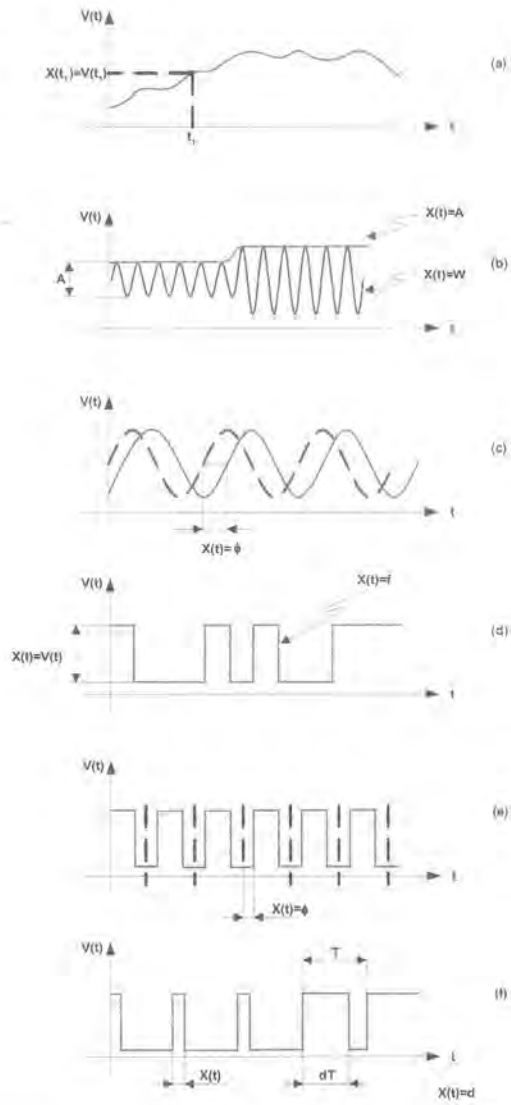


Figura 1.2. Señales que guardan información acerca de variables: el valor que interesa se guarda en (a) el valor instantáneo, (b) la amplitud o (c) la fase de una señal periódica, (d) la frecuencia de una señal digital, (e) la fase de una señal digital o en el ancho de pulso como en (f).

Teniendo en cuenta lo anterior, tenemos que una señal analógica v(t) puede contener información x(t) en cualquiera de los parámetros que la definen:

a) En su **valor instantáneo** en cuyo caso la variable contenida coincide con la señal o es proporcional a ella según aparece en la figura 1.2a:

$$v(t) = Kx(t) \tag{1.1}$$

b) En su **amplitud** en el caso de ser una señal periódica (figura 1.2b):

$$v(t) = K'x(t)\,\text{sen}(\omega t) \tag{1.2}$$

c) En su **frecuencia**, también en el caso de estar en presencia de una señal periódica (figura 1.2b):

$$v(t) = A\,\text{sen}\,(K''x(t)t) \tag{1.3}$$

d) En su **fase** (figura 1.2c), con lo que la variable que tenemos es:

$$v(t) = A\,\text{sen}\,(\omega t + K'''x(t)) \tag{1.4}$$

Cuando la señal v(t) es digital, se puede almacenar también información x(t) tanto digital como analógica en los parámetros que determinan el aspecto de la señal:

a) Información en el **nivel**, en cuyo caso, la información debe ser digital y la señal deberá contener —al menos— tantos niveles como los que se quieran distinguir en la variable original. En la figura 1.2d se ha dibujado una señal binaria que, sólo podrá almacenar dos posibles estados de la variable.

b) En su **frecuencia** (figura 1.2d) lo que constituye un caso particular del de una variable analógica pero en la que se ha sustituido la senoide por una onda cuadrada.

c) En su **fase** (figura 1.2e) también similar al caso anterior de señal analógica.

d) En la **duración de los pulsos** o en el **ciclo de trabajo** (*duty cycle*)[2] de la señal (figura 1.2f).

Hay que tener en cuenta que una determinada señal puede contener más de una variable de información, almacenada en sus diversos parámetros.

Pero, desgraciadamente, tanto las variables como las señales no suelen ser tan simples como las indicadas en las figuras 1.1 y 1.2, sino que contienen factores que tienden a dificultar su lectura. En efecto, cualquier señal o variable puede ser "contaminada" por otras variables no deseadas y de difícil control a lo que llamaremos, en general, "*ruido*". Aunque esto se tratará más exhaustivamente en capítulos posteriores, sí que interesa clasificar las señales desde este punto de vista como:

a) **Señales deterministas** en las que la señal sólo contiene la información o informaciones que interesan (figura 1.3a).

b) **Señales aleatorias o pseudoaleatorias** en las que, sobre los niveles de la señal que interesa, aparecen niveles de ruido (figura 1.3b) que modifican el valor de la señal. Estos niveles pueden afectar o no a la variable (información contenida).

En general, todas las señales tienen un cierto componente de aleatoriedad lo que implica que todas son aleatorias y ninguna es estrictamente determinista; sin embargo, si los niveles del ruido son bajos en relación con los de la señal o su información no se ve afectada por aquel, cabe considerar el caso como determinista y esto es lo que se hace en multitud de ocasiones.

Cuando los niveles de ruido son elevados, la señal cae dentro del área de las aleatorias y requerirá un tratamiento estadístico específico. A este respecto, cabe decir que el ruido que se añade en la señal puede tener dos orígenes distintos:

a) Ruido que no contiene información alguna y que corresponde a la idea exacta de una señal aleatoria que se añade a la nuestra. Se trata de un ruido sin ningún tipo de estructura.

b) Ruido que sí contiene información y que tiene estructura definida, es decir, un patrón de comportamiento y que, por ello, no cabe calificarlo como aleatorio y se usará el término pseudoaleatorio

[2] El *duty cycle* o ciclo de trabajo se define para una señal binaria periódica como el cociente entre el tiempo en que la señal está en el nivel alto y su periodo.

(figura 1.3c). Este ruido puede contener información útil que interese extraer o información no deseada como en el caso de una interferencia de 50 Hz de red que "se cuela" en nuestro sistema: no es aleatorio ya que sigue el patrón de la red eléctrica (con una cierta aleatoriedad) pero no nos interesa para nada. En cualquier caso, dado lo difícil de predecirlo o controlarlo cabe calificarlo como ruido.

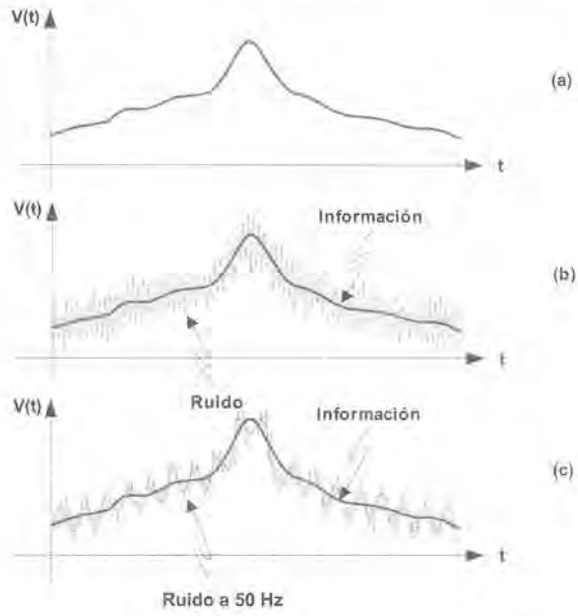


Figura 1.3. (a) Señal determinista; (b) señal aleatoria que incluye información y ruido aleatorio; (c) señal pseudoaleatoria que incorpora la información y un ruido con estructura.

1.1.2 LA INSTRUMENTACIÓN ELECTRÓNICA EN EL CONTROL DE PROCESOS

La instrumentación electrónica cobra su máximo protagonismo en el entorno industrial y adquiere una relevancia especial en lo que hace referencia al **control de procesos**. La realidad es que estas dos disciplinas, instrumentación y control, están fuertemente imbricadas hasta el punto de que comparten algunos criterios de diseño, afectando las decisiones de una directamente a la otra. Para conocer el ámbito y la aplicación de la instrumentación electrónica en este contexto, realizaremos primero una breve introducción que tiene por objetivo familiarizarnos con los conceptos y la terminología relacionada con los sistemas de control.

El concepto de control es extraordinariamente amplio y abarca escenarios de complejidad muy diversa. En su concepción más simple, el control alude al gobierno de un sistema por otro sistema. La figura 1.4 muestra el diagrama de bloques de un sistema de control genérico.

Figura 1.4. Sistema de control genérico.

El objetivo de un sistema de control es obtener una salida, es decir, una respuesta que coincida con la que pretende el operador pero sin que intervenga directamente sobre el sistema. Para ello, el operador se sirve de unas entradas, denominadas **consignas** o **variables de control**, que le permiten especificar la respuesta deseada de la planta. A partir de estas señales, el sistema de control genera las denominadas **señales de mando** que son las que actúan sobre la planta con objeto de modificar la salida del proceso. Generalmente, el sistema de control funciona con magnitudes de baja potencia, llamadas **señales**, y gobierna unos **accionamientos** o **actuadores**. Esta idea se refleja en la figura 1.5 donde el conjunto formado por el sistema de control y los accionamientos ejecutan las órdenes dadas por el operador a través de las entradas de consigna.

Figura 1.5. Sistema de control en lazo abierto.

Este tipo de sistemas se denomina en "**lazo abierto**", debido a que el sistema de control no recibe ningún tipo de información del comportamiento de la planta, es decir, la salida del proceso no afecta a la acción de control.

Figura 1.6. Sistema de control realimentado o en lazo cerrado.

En un sistema de control en lazo abierto, a cada consigna le corresponde un modo de funcionamiento fijo. Este método se utiliza cuando se conoce de antemano la relación entre la entrada y la salida y no existen perturbaciones de ningún tipo. Una **perturbación** es una señal que modifica negativamente la salida de un sistema. Si sobre un sistema se producen perturbaciones impredecibles, los sistemas de control en bucle abierto no pueden corregir la situación ya que el sistema de control no recibe información sobre ese hecho. Resulta interesante la consideración de "perturbación impredecible" ya que si fuesen predecibles siempre se podrían compensar dentro del sistema. Cuando se producen perturbaciones impredecibles y pretendemos que la planta se comporte de la forma deseada, es necesario informar al sistema de control sobre esta situación. La figura 1.6 muestra el diagrama de bloques genérico de un sistema de control realimentado. Ahora, el operador fija las variables denominadas de **consigna** o de **referencia**, que establecen el comportamiento deseado, y el sistema de control genera las **señales de mando** adecuadas para conseguir que la salida del

proceso, **variable regulada**, se mantenga en el valor deseado a pesar de las perturbaciones exteriores. **Regular** una variable significa mantener esa variable en un valor deseado frente a las influencias externas. Este tipo de sistemas se denominan "**sistemas de control en lazo cerrado**" o "**realimentados**". En esta configuración se alimenta al controlador con la denominada "**señal de error**" que es la diferencia entre la señal de referencia y la señal de realimentación. La señal de realimentación puede ser la propia señal de salida, variable regulada, o una función de ésta. Una ventaja de los sistemas de control realimentados es que desensibiliza el sistema frente a las perturbaciones exteriores e incluso frente a las variaciones internas de los parámetros del sistema.

EJEMPLO 1.1

El siguiente ejemplo trata de ilustrar estos conceptos mediante un sistema de control en lazo cerrado basado en una piscina cuyo nivel de llenado se pretende controlar. El sistema consta de una válvula para el llenado de la piscina, otra para el vaciado y un flotador para conocer el nivel actual del líquido. El controlador se encarga de mantener el nivel del líquido a un valor deseado, especificado en la entrada de consigna de la figura 1.7a a pesar de las inclemencias meteorológicas.

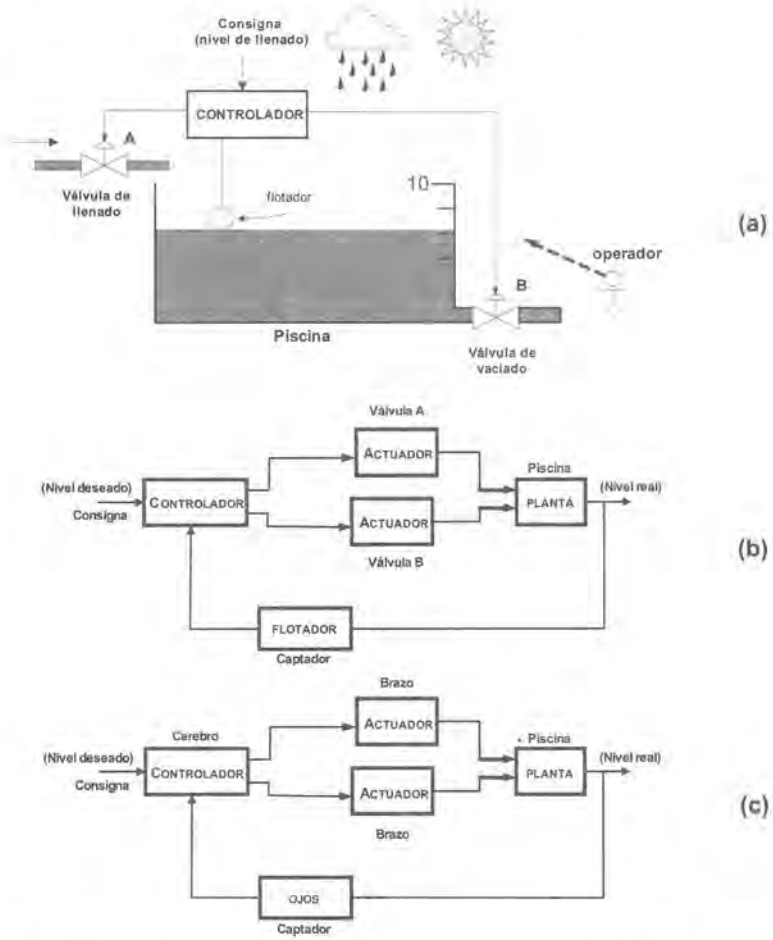


Figura 1.7. Sistema de control en lazo cerrado. (a) Sistema de control del nivel de agua en un depósito; (b) diagrama de bloques del sistema de control con realimentación automática; (c) diagrama de bloques del sistema de control con realimentación manual.

La comparación entre el nivel real del depósito, variable regulada, y el nivel deseado, consigna, genera una señal de error. En el caso de que el nivel sea inferior al establecido, el sistema de control genera una señal de mando que actúa sobre la electro-válvula A (el actuador) permitiendo el llenado del tanque (la planta); si fuese superior, se actuaría sobre la electro-válvula B. En caso contrario, la válvula se mantiene cerrada. Para conocer el estado de la planta, es decir, el nivel actual del líquido, se utiliza un flotador ***(el captador o sensor)*** *que informa al sistema de control sobre la variable regulada.*

En la figura 1.7b se representa el diagrama de bloques del sistema descrito, un sistema de control automático en bucle cerrado. En la figura 1.7c se representa el diagrama de bloques en el caso de control manual, lo que equivale a sustituir el bloque de control por un operador. En este caso, el cerebro del operador actúa como sistema de control, sus ojos, el sensor, permiten determinar el valor de la variable regulada, nivel del depósito, y sus brazos, los actuadores, abren o cierran las válvulas de llenado y vaciado en el caso necesario.

Cuando a una planta se le aplica un control automático realimentado, como en el ejemplo descrito, se habla de **control de procesos**. En el contexto del control de procesos, el interés de la instrumentación electrónica radica en las técnicas y procedimientos que esta disciplina aporta para medir la variable de proceso e informar al sistema de control. No se puede controlar una variable de forma más precisa a como la podemos medir, así que la medida es un elemento fundamental para el control. En definitiva, el conocimiento sobre el estado de un proceso se realiza empíricamente a través de la medida, y esta labor puede realizarse recurriendo a procedimientos electrónicos de instrumentación diseñando un **sistema de medida** integrado en el ciclo de control. Cuando se habla de instrumentar cualquier sistema físico se refiere a añadirle todos los sistemas de captación que sean capaces de leer los parámetros físicos que formen parte de él. La figura 1.8 representa los elementos esenciales de un sistema de medida en el contexto del control de procesos.

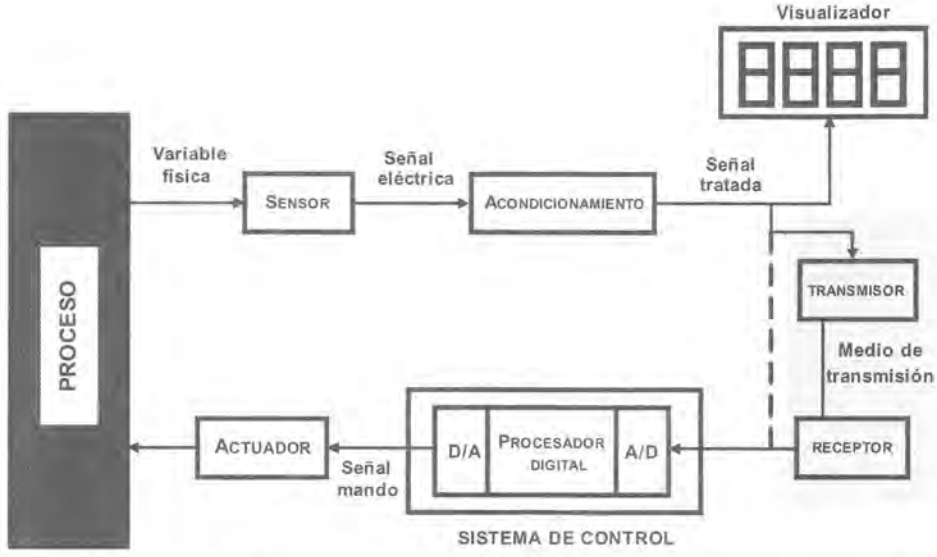


Figura 1.8. Diagrama de bloques de un sistema de medida en el contexto del control de procesos.

Los bloques básicos que se reflejan en la figura, y que estudiaremos con más detalle en la siguiente sección, son:

- **Sensor:** su misión es capturar la variable de proceso, variable física, y convertirla en una señal eléctrica.

- **Acondicionamiento de señal:** trata la señal eléctrica para adaptarla al siguiente bloque de la cadena. El acondicionamiento puede incluir:
  - Amplificación.
  - Filtrado.
  - Conversión de niveles.
  - Conversión de tensión a corriente, y viceversa.
  - Conversión de tensión a frecuencia, y viceversa.
  - Linealización, etc.
- **Visualización:** unidad que presenta la información al operador.
- **Sistema de control:** procesa los datos de acuerdo al algoritmo de control y genera la señal de mando. Los sistemas de control pueden ser:
  - Analógicos: emplean circuitos analógicos para el control.
  - Digitales: basados en procesadores digitales.
- **Actuadores:** conjunto de dispositivos que modifican la respuesta del sistema.
- **Sistemas de transmisión remota:** permite transmitir la información entre la planta y el sistema de control en el caso de que estén alejados uno del otro.

La instrumentación electrónica se ocupa fundamentalmente de los bloques relacionados con la captura de señales de proceso y el acondicionamiento de señal aunque su ámbito de estudio también incluye los actuadores, el diseño de los sistemas reguladores y comparadores, el diseño de los sistemas de transmisión, la representación o visualización de magnitudes, etc. En realidad, la frontera entre instrumentación y control no siempre es evidente, aunque el ámbito del control se delimita cuando los criterios de diseño quedan establecidos por la teoría de regulación automática. Algunos autores solventan el problema hablando de instrumentación electrónica de medida e instrumentación electrónica de control, haciendo referencia, respectivamente, a la captación (incluyendo los procesos de acondicionamiento) y al accionamiento (incluyendo los sistemas de toma de decisiones). En este texto se hará referencia a la instrumentación electrónica de medida a la que denominaremos sencillamente, instrumentación electrónica.

1.2 Sistemas de medida

Un sistema de medida electrónico es aquel equipo cuya finalidad es obtener información acerca de un proceso físico y presentar dicha información en la forma adecuada a un observador o a otro sistema técnico de control (figura 1.9). Existen multitud de magnitudes físicas susceptibles de ser medidas, si bien cada una de ellas podría considerarse dentro de los tipos de variables que muestra la tabla 1.1.

Figura 1.9. Objeto de un sistema de medida.

Naturaleza de la variable	Tipo de variable
Mecánica	Desplazamiento, velocidad, aceleración, fuerza, par, presión, masa, flujo, etc.
Térmica	Temperatura, calor, entropía, etc.
Magnética	Campo magnético, flujo, permeabilidad magnética, etc.
Eléctrica	Carga, corriente, tensión, resistencia, conductancia, capacidad, permitividad dieléctrica, polarización, frecuencia, etc.
Óptica	Rayos gamma, rayos X, ultravioleta, visible, infrarrojo, microondas, etc.
Química	Humedad, pH, concentración iónica, análisis de gases, etc.
Biológica	Proteínas, hormonas, antígenos, etc.

Tabla 1.1. Naturaleza y tipos de variables.

1.2.1 FUNCIONES DE UN SISTEMA DE MEDIDA

Considerando el sistema de medida como una caja negra, la entrada sería el **valor verdadero** de la variable a medir y la salida, el **valor medido**. Sólo en el caso ideal, la diferencia entre ambos valores será nula por lo que siempre cometeremos un **error de medida**. Las causas de este error son de origen diverso: ruido del sistema de medida, interferencias exteriores, desviaciones en los parámetros de componentes, mala calibración, etc. En un sistema de medida podemos distinguir tres funciones principales: adquisición de datos, procesamiento de datos y distribución de los datos (figura 1.10):

- **Adquisición de datos**. La información de las variables a medir es adquirida y convertida en una señal eléctrica. De esta etapa dependerá en gran medida las prestaciones del sistema de medida.
- **Procesamiento de datos**. Consiste en el procesamiento, selección y manipulación de los datos con arreglo a los objetivos perseguidos. Esta función suele ser realizada por un procesador digital, tipo microcontrolador o procesador digital de señal (*Digital Signal Processor* o DSP).
- **Distribución de datos**. El valor medido se presenta a un observador (por ejemplo, mediante un display), se almacena (por ejemplo, en disco o en un chip de memoria) o bien se transmite a otro sistema.

Figura 1.10. Funciones principales de un sistema de medida electrónico.

A su vez, la función de adquisición de datos puede ser dividida en unidades funcionales más pequeñas como se muestra en la figura 1.11. En primer lugar, la variable del mundo físico es convertida en una señal eléctrica mediante un dispositivo **sensor** a fin de poder ser procesada adecuadamente. Con frecuencia, la señal procedente del sensor tiene unas características que la hacen poco adecuada para ser procesada: señal de pequeño nivel, espectro grande, falta de linealidad, etc. Se hace, pues, necesaria una etapa de **acondicionamiento** de la señal. Este acondicionamiento consiste en realizar alguna de las siguientes operaciones básicas:

- **Amplificación:** incrementar el nivel de potencia de la señal.
- **Filtrado:** eliminar las componentes de la señal no deseadas.

- **Linealización:** obtener una señal de salida que varíe linealmente con la variable que se desea medir.
- **Modulación/Demodulación:** modificar la forma de la señal a fin de poder transmitirla a largas distancias o a fin de reducir su sensibilidad frente a interferencias durante el transporte.

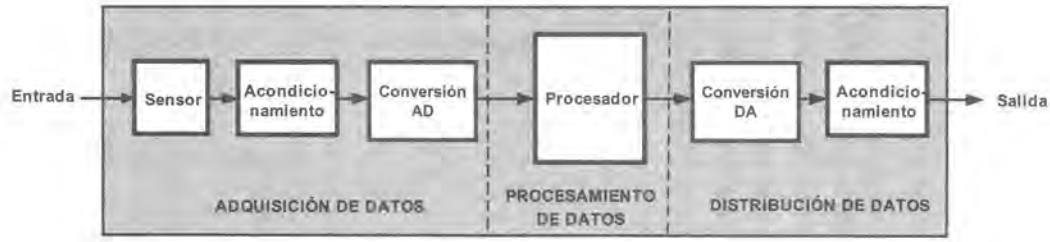


Figura 1.11. Elementos de un sistema de medida.

En un esfuerzo por simplificar el diseño, ahorrar espacio, eliminar fuentes de errores y reducir costes, los sensores suelen incluir circuitería de acondicionamiento, dando lugar a sensores integrados, que incluso proporcionan una salida digital. Después del acondicionamiento, la señal es convertida del mundo analógico al dominio digital mediante un convertidor analógico-digital (ADC). La salida del ADC se procesa mediante un procesador digital de señales, que en muchas ocasiones puede integrar el ADC. Una vez procesada la señal puede ser necesario entregar la información en forma analógica, en cuyo caso es necesario proporcionar otra interfaz desde el dominio digital al mundo analógico mediante un convertidor digital-analógico (DAC). Finalmente, la señal del procesador se puede enviar a un observador para su visualización, almacenada en memoria o enviar a otro equipo de medida o sistema de control.

Procesar digitalmente una señal tiene sus ventajas, que, aunque las iremos descubriendo a lo largo del texto, conviene ya insinuar: flexibilidad de diseño (en el sentido de poder variar por software la operación), mayor precisión que con los circuitos analógicos, las señales se almacenan más fácilmente en memoria, etc. Estas ventajas no implican que el procesado de señal digital sea siempre la solución a todos problemas de procesamiento. Así, señales de gran ancho de banda que puedan requerir un procesamiento en tiempo real precisan un procesado analógico.

1.2.2 SISTEMAS DE MEDIDA MULTICANAL

El esquema de la figura 1.11 corresponde a un sistema de un solo canal de medida. En la mayoría de las situaciones se hace necesario procesar más de una variable de entrada. La figura 1.12 muestra una primera solución, en la que un multiplexor analógico se encarga de llevar las señales acondicionadas de los sensores a un único ADC.

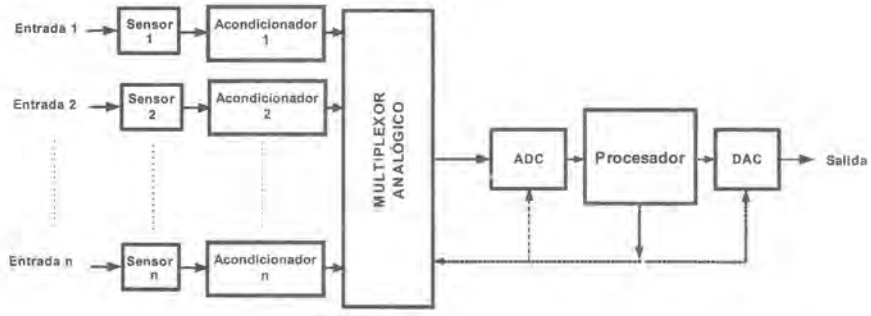


Figura 1.12. Sistema de medida multicanal con un solo ADC.

La estructura anterior no es adecuada cuando la aplicación requiere una captura simultánea de las variables (por ejemplo, la medida de un desfase entre dos señales). En este caso, se recurre a un sistema de medida con conversión en paralelo, como el mostrado en la figura 1.13, en el que cada canal de entrada dispone de un convertidor A/D independiente. Otro caso habitual es el de las tarjetas de adquisición de datos, en las cuales suele aprovecharse el propio procesador del computador para las tareas de procesamiento.

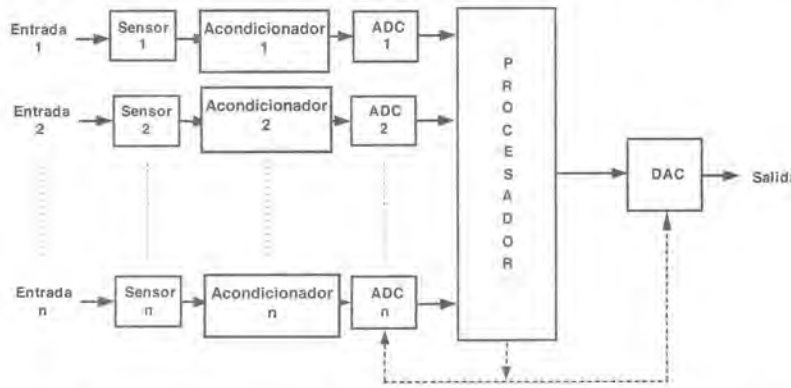


Figura 1.13. Sistema de medida multicanal con un ADC por canal.

EJEMPLO 1.2: Sistema de instrumentación industrial

La medida de la temperatura es, sin duda, una de las mediciones más frecuentes y más importantes que se suelen efectuar en los procesos industriales. La mayoría de los sistemas físicos, electrónicos, químicos, mecánicos y biológicos se ven afectados de una u otra forma por la temperatura. Algunos procesos trabajan bien sólo dentro de un margen estrecho de temperatura; ciertas reacciones químicas, procesos biológicos e incluso los circuitos electrónicos se comportan mejor dentro de ciertos límites de temperatura. Cuando estos procesos necesitan ser optimizados, se requiere la medida de la temperatura. Los sensores de temperatura son la entrada a estos sistemas de control.

La figura siguiente muestra un sistema de medida de la temperatura de un horno industrial. Esta variable ha de estar comprendida entre ciertos límites. La señal proporcionada por el sensor de temperatura es acondicionada en las proximidades del horno y enviada a un dispositivo de control que puede o no encontrarse en el mismo lugar del horno. En este caso, la señal puede enviarse directamente (mediante bucle de

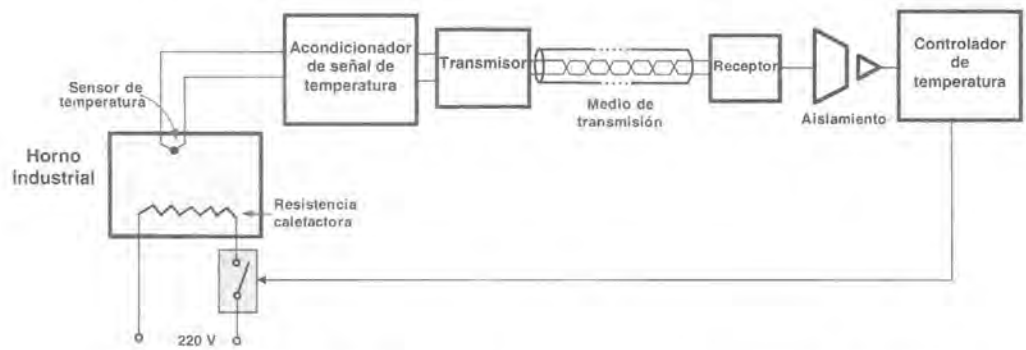


Figura 1.14. Ejemplo de sistema de instrumentación de aplicación industrial.

corriente) o bien modulada. El controlador, en función del valor medido de la temperatura, dará las órdenes oportunas al dispositivo de potencia para mantener la temperatura dentro de los límites requeridos.

EJEMPLO 1.3: Sistema de instrumentación biomédica

La instrumentación biomédica se encarga de la medida de las señales procedentes de un ser vivo. Las señales a medir proceden de situaciones clínicas diferentes: diagnóstico, terapia, cirugía, rehabilitación, etc. Estas señales han de ser procesadas garantizando en todo momento la seguridad del paciente, de ahí que muchas veces se recurra a métodos de medida indirectos.

La figura siguiente muestra un esquema básico de un sistema para monitorizar la actividad eléctrica del corazón (electrocardiograma-ECG). Para ello, se sitúan una serie de electrodos en distintas partes del cuerpo obteniéndose unas señales de amplitud y frecuencia características. Las señales obtenidas son amplificadas mediante un amplificador de instrumentación. Como ya se ha comentado, la seguridad del paciente es vital en este tipo de aplicaciones, por ello se introduce un aislamiento eléctrico entre el sistema de medida y el paciente. Dentro del acondicionado de la señal es necesario realizar un filtrado de la misma a fin de limitar el ancho de banda de las señales de interés al estrictamente necesario. Todos estos bloques serán estudiados a lo largo del libro, por lo que no cunda el pánico.

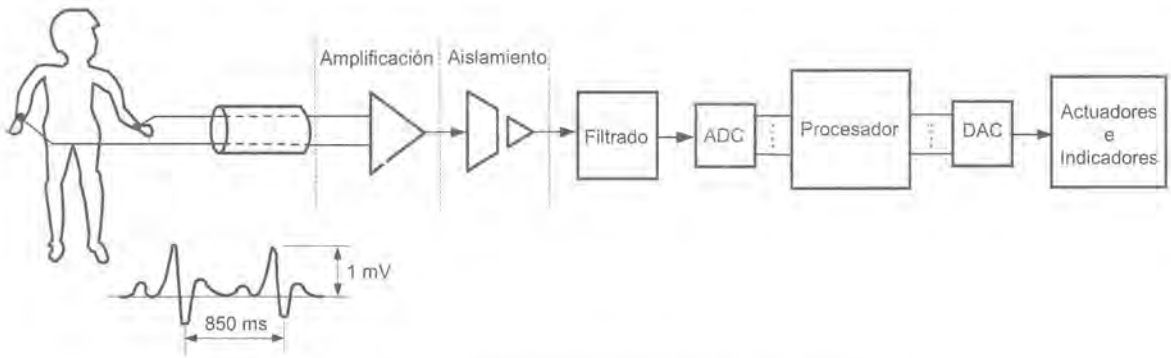


Figura 1.15. Ejemplo de sistema de instrumentación de aplicación médica.

1.2.3 ARQUITECTURAS DE LOS SISTEMAS DE INSTRUMENTACIÓN

En los procesos tecnológicos son muchas las variables de las que hay que tener información oportuna para mantenerles en un punto de operación determinado. En tales casos, el sistema de medida y control puede adoptar dos arquitecturas básicas:

Arquitectura centralizada

La figura 1.16 presenta esta arquitectura que se caracteriza por realizar la ejecución del algoritmo de medición y control en un núcleo inteligente, es decir, centralizadamente. Generalmente, se emplea en procesos de pocas variables y con distancias cortas entre los sensores y el núcleo inteligente. Sus desventajas son la necesidad de mucho cable generalmente de alto coste y las elevadas exigencias sobre el sistema de acondicionamiento de señales debido al ruido eléctrico presente.

Arquitectura distribuida

La figura 1.17 presenta el esquema general de esta arquitectura. Es utilizada cuando el número de señales del proceso es muy elevado, o su dispersión geográfica es muy grande y/o cuando las exigencias dinámicas de las variables medidas son altas. Se caracteriza por poseer varios núcleos inteligentes (1, 2, 3...N), sobre la base de microprocesadores, microcontroladores, que se comunican con otros sistemas a través de un bus de proceso digital (generalmente a 2 hilos), por el que fluye la información en forma serie y con alta inmunidad al ruido.

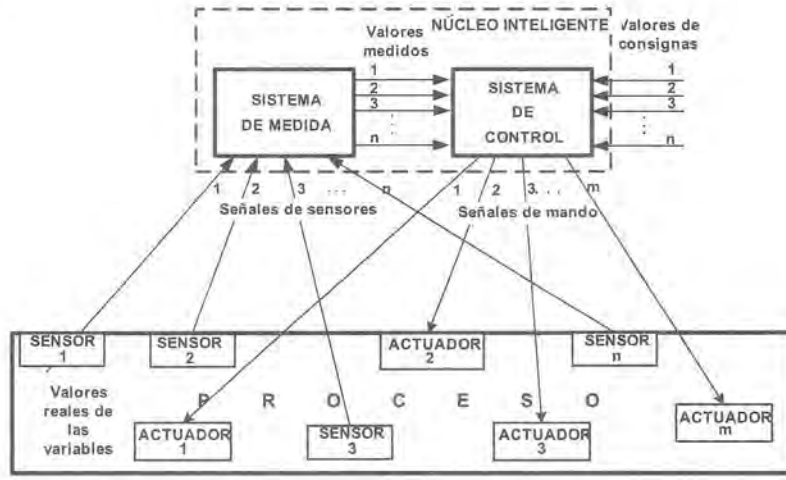


Figura 1.16. Arquitectura centralizada.

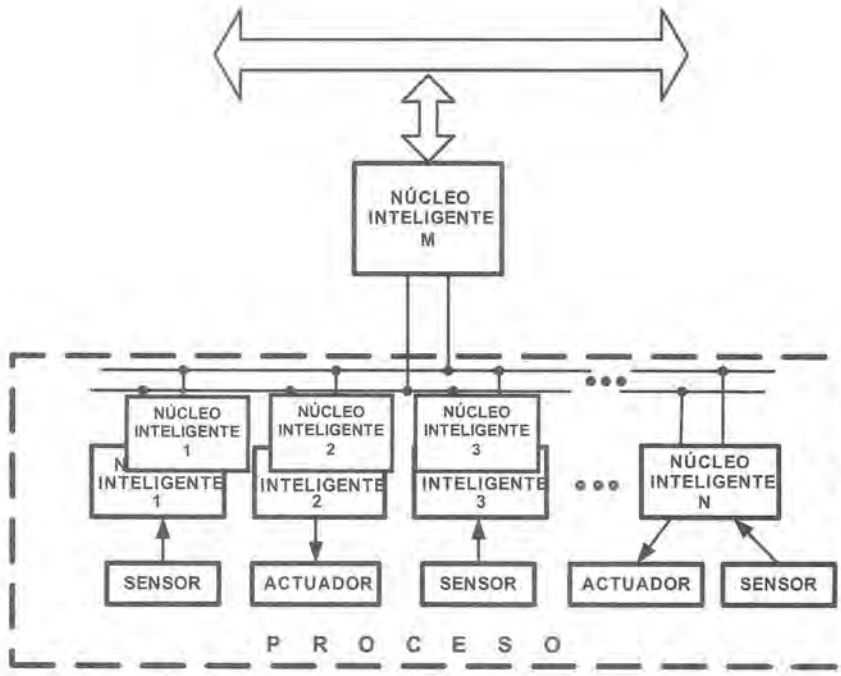


Figura 1.17. Arquitectura distribuida.

El núcleo inteligente de segundo nivel (M) realiza las funciones de organización de la transferencia de información y ejecuta parcial o totalmente el algoritmo de medición y control. Pueden existir varios de estos sistemas que intercambien información digital con mayor rapidez e inmunidad al ruido a través de una estructura en bus. A diferencia de la arquitectura de estrella, en esta estructura se realiza el procesamiento de la información **descentralizadamente.** El núcleo ejecutor del algoritmo de medición y control posee entradas y salidas digitales, a través de los buses de proceso y de campo, respectivamente. Al propio tiempo, cada terminal de medición (1, 2, 3...N) se encuentra muy próximo a los sensores o en el propio sensor (sensores inteligentes o *smart sensors*) y en ellas se realiza el procesamiento primario de la información (adquisición, filtrado, validación, linealización, calibración, etc.).

1.3 Características estáticas

El comportamiento de un sensor o de un instrumento de medida se puede definir, en general, mediante la función de transferencia, que indica tanto el comportamiento en régimen estático como dinámico. El primero corresponde a la relación entre la entrada y la salida cuando la entrada es constante o cuando ha transcurrido un tiempo suficiente para que la salida haya alcanzado el valor final o régimen permanente. El segundo indica la evolución del sistema hasta que la salida alcanza el valor final ante una variación en la entrada.

Una función de transferencia que recogiese con rigurosidad ambos comportamientos resultaría tremendamente compleja por lo que, en la práctica, suelen indicarse por separado mediante una serie de parámetros. El presente apartado define las características estáticas.

1.3.1 CURVA DE CALIBRACIÓN (*STATIC TRANSFER FUNCTION*)

La curva de calibración es la relación entre la entrada al sensor o sistema y su salida en régimen estático (figura 1.18).

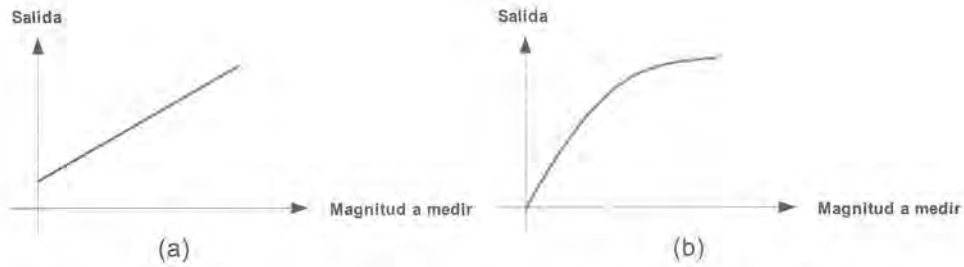


Figura 1.18. Ejemplos de curva de calibración. (a) Curva de calibración lineal. (b) Curva de calibración no lineal.

Para definir la curva adecuadamente se necesita como mínimo indicar su forma y sus límites. Estos últimos se especifican con algunos de los siguientes parámetros (figura 1.19):

- **Campo de medida** (*range*): conjunto de valores comprendidos entre los límites superior e inferior entre los que puede efectuarse la medida. Por ejemplo, si se dispone de un termómetro diseñado para medir entre –20 y 60°C, el campo de medida es de –20°C/60°C. También se conoce como rango dinámico aunque este término se emplea más frecuentemente para indicar la diferencia entre el límite superior e inferior expresada en decibelios.

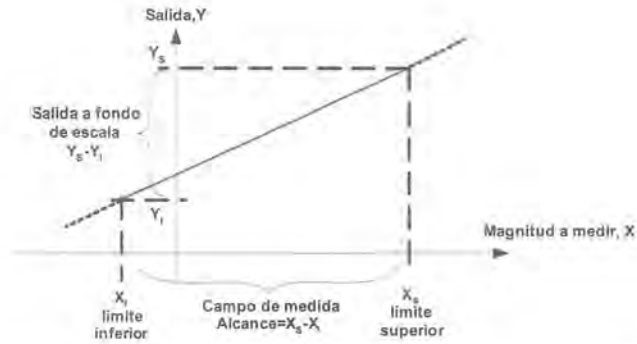


Figura 1.19. Representación del campo de medida, alcance y salida a fondo de escala sobre la curva de calibración.

- **Alcance, fondo de escala** (*span, input full scale —FS—*): es la diferencia entre los límites superior e inferior de medida. Por ejemplo, en el caso del termómetro mencionado anteriormente el alcance es de 80°C. También se conoce con el término *fondo de escala*. Conviene no confundir este término con el límite superior de medida, que es de 60°C, aunque en muchos sistemas el alcance y el límite superior puedan coincidir.
- **Salida a fondo de escala** (*full scale output, —FSO—*): es la diferencia entre las salidas para los extremos del campo de medida.

En cuanto a la forma de la curva de la calibración, hay que tener en cuenta que muchos de los sensores de mayor interés presentan una respuesta que puede aproximarse a una línea recta. Por tanto, la curva de calibración puede definirse fácilmente mediante dos puntos o mediante un punto y la pendiente, siendo esta última definición mucho más habitual. No obstante, cuando la recta pasa por el origen sólo es necesario especificar la pendiente. Además, es necesario indicar, de algún modo, el error cometido en la aproximación, es decir, la diferencia entre la curva real y la curva linealizada. Por ello, para definir la curva linealizada se emplean los siguientes términos:

- **Sensibilidad** (*sensitivity*): es la pendiente de la curva de calibración.
- **No linealidad** (*nonlinearity*): es la máxima desviación de la curva de calibración con respecto a la línea recta por la que se ha aproximado. Habitualmente se suele expresar en % con respecto al alcance y suele denominarse como error de linealidad e incluso como linealidad. Téngase en cuenta que este error depende de la aproximación que se haya realizado. Así, por ejemplo, en la figura 1.20 se muestra la misma curva de calibración linealizada por dos formas diferentes siendo el error también diferente.

  En muchos casos, la linealización se suele realizar atendiendo a los extremos del campo de medida o atendiendo a la zona más habitual de medida. Por ejemplo, si se pretende realizar un termómetro para aplicaciones domésticas puede resultar conveniente utilizar como uno de los puntos de linealización el que corresponde a la temperatura ambiente habitual (por ejemplo, 20°C) de forma que el error alrededor de esta zona sea mínimo. También se puede realizar por el método de mínimos cuadrados con lo que el error será el mínimo posible sobre el conjunto del campo de medida [1.9], aunque en la práctica la calibración del sistema resulta difícil de llevar a cabo. Este último método de linealización suele ser el que emplean los fabricantes de los sensores para indicar el error de linealidad de los mismos, también llamado linealidad independiente.

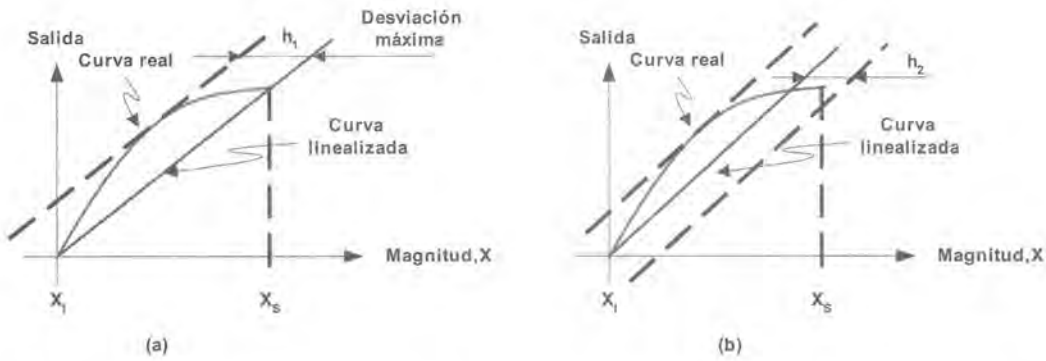


Figura 1.20. Ejemplos de linealización de una curva de calibración. (a) El ajuste de la recta se ha realizado mediante los extremos de medida; la no-linealidad es igual a $h_1 \cdot 100/(X_S - X_I)$ [%]. (b) Otra posible linealización que lleva a una menor desviación; no-linealidad igual a $h_2 \cdot 100/(X_S - X_I)$ [%]

EJERCICIO 1.1

Se dispone de un sensor cuya salida, "y" en función de la magnitud a medir, "x" es:

$$y_r(x) = \frac{5x}{100 + x}$$

donde $0 \leq x \leq 10$.

Si se linealiza la respuesta por la función:

$$y_l(x) = \frac{5x}{100}$$

¿Cuál es el error de linealidad cometido?

SOLUCIÓN:

Como se puede observar en la figura 1.21, en la que se han representado ambas funciones, la máxima desviación de la respuesta linealizada con respecto a la real, se encuentra para x = 10.

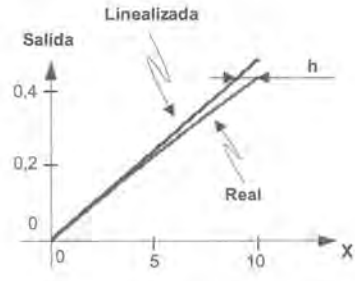


Figura 1.21. Representación gráfica de la respuesta real y de la respuesta linealizada. Se puede observar que el máximo error tiene lugar en el extremo superior de medida.

Para x = 10, la respuesta real es $y_r(10) = 0{,}46$. Sin embargo, esta respuesta corresponde para la curva linealizada a un valor x = 9,09. Por lo tanto, la desviación h, en valor absoluto es h = 0,91. Así pues, el error, expresado en % sobre el fondo de escala es:

$$\frac{h}{10}\cdot 100 = 9{,}1\%$$

Para las curvas de calibración que no son lineales, la sensibilidad resulta insuficiente para definirlas puesto que es variable a lo largo de la misma; no obstante, suele resultar de interés conocer las sensibilidades máxima y mínima. Generalmente las curvas se definen mediante una tabla indicando para cada entrada la correspondiente salida. También resulta muy habitual aproximar la salida a alguna función matemática, siendo necesario especificar, de nuevo, el error cometido en tal aproximación.

En algunos casos, la definición de la curva de calibración puede requerir algún parámetro adicional, entre los que cabe destacar los siguientes:

- **Zona muerta** (*dead zone*): es el campo de valores de la variable que no hace variar la indicación. También se define como la región de la curva de calibración que presenta una sensibilidad nula. Por uejemplo, en los potenciómetros angulares de una vuelta sin fin, la variación de la resistencia entre el cursor y uno de los terminales con el ángulo de giro es, generalmente, lineal; pero existen ciertos ángulos de giro que no producen variación alguna (figura 1.22).

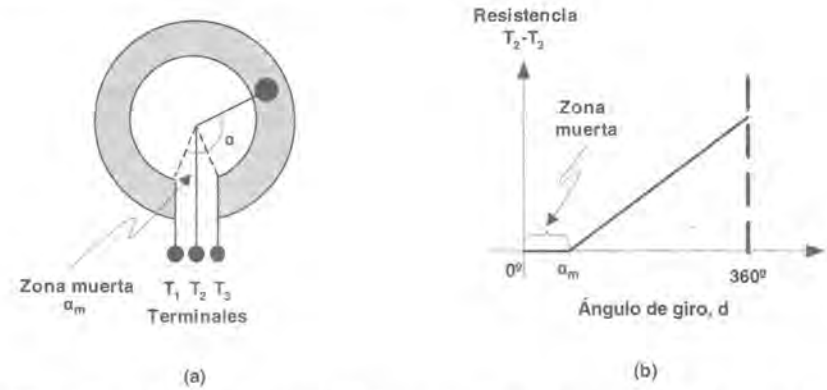


Figura 1.22. Ejemplo de zona muerta. (a) Esquema de un potenciómetro de una vuelta sin fin en el que se puede apreciar la zona muerta o ángulo de giro, α_m, que no proporciona una variación en la resistencia entre los terminales T_2 y T_3. (b) Curva de calibración correspondiente.

- **Histéresis** (*hysteresis*): es la diferencia en la medida dependiendo del sentido en el que se ha alcanzado (figura 1.23). Por ejemplo, si se dispone de una báscula cuyo campo de medida es 0-100 kg y se va incrementando el peso sobre la misma gradualmente hasta llegar al máximo, es probable que la indicación del peso no sea la misma que si se va disminuyendo desde el máximo. Si ante 50 kg, en el primer caso marca 49 kg y en el segundo 51 kg, la histéresis es de 2 kg o bien, expresada respecto del fondo de escala, es del 2%.

- **Deriva** (*drift*): es la variación de algún aspecto de la curva de calibración con respecto a algún parámetro ambiental (temperatura, humedad, etc., siempre que el propio parámetro no sea el objeto de la medida) o con respecto al tiempo. Para caracterizar los sistemas suele ser muy habitual indicar la deriva en algún punto significativo como, por ejemplo, el que corresponde al cero. Normalmente se expresa en % sobre el fondo de escala de salida.

- **Saturación** (*saturation*): es el nivel de entrada a partir del cual la sensibilidad disminuye de forma significativa (figura 1.24).

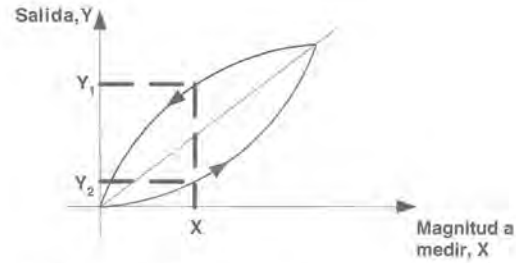


Figura 1.23. Curva de calibración que presenta histéresis.

- **Resolución** (*resolution*): en algunos sensores e instrumentos de medida, una variación continua de la señal de entrada no da lugar a una variación continua en la salida, más bien parece tener una forma escalonada. La resolución se define como el incremento mínimo de la variable de entrada que ofrece un cambio medible a la salida. Es un parámetro también muy importante, por ejemplo, en los convertidores analógico a digital o en aquellos equipos de medida que utilicen estos convertidores como paso previo a la visualización de los resultados (figura 1.25). Habitualmente, se suele expresar como un valor en tanto por ciento sobre el fondo de escala. Por ejemplo, en un potenciómetro bobinado de 100 espiras, la resolución es del 1%.

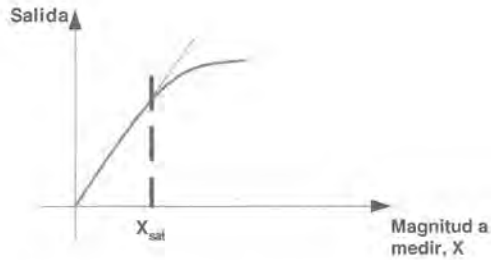


Figura 1.24. Curva de calibración que comienza a saturarse por encima de X_{Sal}.

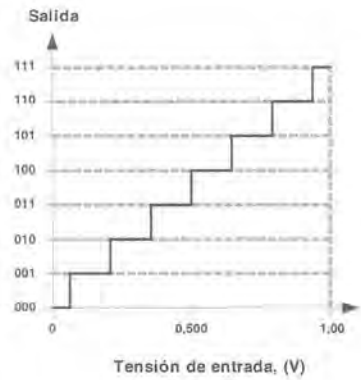


Figura 1.25. Relación entre la salida y la tensión de entrada de un convertidor analógico a digital de 3 bits de 0-1 V.

1.3.2 ERRORES: EXACTITUD, VERACIDAD Y PRECISIÓN

Un principio básico de todo sistema de instrumentación electrónica es el de medir una magnitud con el menor error posible. Siempre existe un grado de incertidumbre puesto que es imposible realizar una medición sin modificar en mayor o menor grado aquello que se mide[3]. Por suerte, no es necesario conocer las magnitudes con exactitud absoluta. Vivimos en un mundo de aproximaciones (si pudiéramos cuantificar nuestro peso con infinitos decimales, nos moriríamos antes de poder leerlo) y nos resulta suficiente con conocer las magnitudes con un cierto grado de error. Por ejemplo, en el epígrafe anterior ya se ha realizado la aproximación a una curva de calibración (ya aproximada, nunca será exacta) mediante una línea recta siempre que el error cometido sea aceptable en nuestra aplicación. Además, las variables incontroladas, entre ellas el ambiente, el envejecimiento de los componentes, el ruido, etc., añaden nuevos errores.

Los términos más elementales para cuantificar el error son el error absoluto y el error relativo:

- **Error absoluto:** diferencia entre el valor medido y el valor exacto en valor absoluto.
- **Error relativo:** error absoluto dividido entre el valor exacto.

Por otra parte, los errores se suelen clasificar en sistemáticos y aleatorios. Los primeros se mantienen constantes en valor absoluto y signo después de medidas repetidas por lo que pueden ser evitados, mientras que los segundos permanecen aún después de haber corregido los primeros [1.16]. Ambos son, en principio, independientes por lo que si el error sistemático en la medida de una magnitud x es $\pm k_1 x$ y el aleatorio es $\pm k_2 x$, el error total es [1.14]:

$$e = \sqrt{(k_1 x)^2 + (k_2 x)^2} \tag{1.5}$$

Una mejor definición del comportamiento del sistema en lo referente al error exige cuantificar el error de forma que pueda detectarse el origen del mismo. Así, en las normas UNE82009 [1.1-1.6] se definen los siguientes términos para cuantificar el error.

- **Veracidad** (*trueness*): es el grado de concordancia entre el valor medio obtenido de una gran serie de resultados y el valor verdadero o el aceptado como referencia. En muchos métodos de medición no se conoce el valor verdadero, pero es posible contar un valor aceptado para la propiedad que se va a medir; por ejemplo, acudiendo a determinados materiales de referencia, estableciendo dicho valor por referencia a otro método de medición o mediante la preparación de una muestra conocida. La veracidad se suele expresar en términos de **sesgo** o **desviación** (*bias*).
- **Precisión** (*precision*): es el grado de concordancia entre los resultados. Observe que la precisión no implica necesariamente la veracidad. Los resultados de muchas medidas pueden ser muy similares pero, sin embargo, pueden ser muy diferentes del valor verdadero.
- **Exactitud** (*accuracy*): este término se utiliza para referirse conjuntamente a la veracidad y a la precisión, es decir, a la correspondencia de los resultados entre sí y, además, al valor verdadero. Aunque es habitual denominar como precisión a la exactitud, las diferencias son notables.

El término general para definir la variabilidad entre mediciones repetidas es la precisión. Existen diferentes factores que pueden contribuir a la variablidad de un método de medición, entre ellos, los siguientes:

- El operador que realiza la medición.

[3] Hasta la llegada de la mecánica cuántica se pensaba que la exactitud de una medida estaba únicamente limitada por la exactitud del instrumento empleado para realizarla. En 1927 el físico W. Heisenberg demostró que no se podía determinar en un instante dado la posición y el momento de un electrón por muy exactos que fueran los instrumentos empleados y, por tanto, no se podría determinar con exactitud cuál sería el movimiento o la posición en un instante posterior, salvo en términos de probabilidad.
El principio de incertidumbre de Heisenberg ha tenido importantísimas implicaciones filosóficas que llegan hasta nuestros días.

- Los equipos.
- La calibración de los equipos.
- El ambiente (temperatura, humedad, etc.).
- El intervalo temporal entre las mediciones.

La precisión se cuantifica a partir de dos términos denominados **repetibilidad** (*repeatability*) y **reproducibilidad** (*reproducibility*). La medición de la repetibilidad se realiza manteniendo constantes los factores anteriores mientras que, bajo condiciones de reproducibilidad, se varían.

EJERCICIO 1.2

Una persona que pesa 70,0 kg, se ha pesado en dos básculas diferentes. Después de pesarse cinco veces con cada una de las básculas los resultados son los siguientes:

Primera báscula: *65,0 kg 65,1 kg 64,9 kg 65,1 kg 65,0 kg.*
Segunda báscula: *70,0 kg 72,0 kg 68,0 kg 71,1 kg 73,0 kg.*

¿Cuál presenta una mayor precisión?

SOLUCIÓN:

Como se puede observar la primera de las básculas ofrece unos valores con una buena concordancia en comparación con la segunda, mientras que esta ofrece siempre unos valores más cercanos al verdadero que la primera, luego la primera es más precisa pero ofrece resultados menos veraces que la segunda.

Se puede observar que la primera báscula parece presentar un error sistemático de alrededor de 5,0 kg. Este error se podría eliminar sumando siempre 5 kg al resultado de la medida. No obstante, seguiría existiendo un error impredecible de 0,1 kg que se puede considerar como aleatorio.

La cuantificación de la exactitud, precisión, veracidad, repetibilidad y reproducibilidad se obtiene y se expresa mediante procedimientos y términos estadísticos [1.1-1.6]. Sin embargo, los fabricantes de instrumentos de medida suelen especificar estos aspectos (principalmente, la exactitud) en % sobre el fondo de escala de entrada o de salida, para que sea un valor fácilmente comprensible y manejable por el personal poco familiarizado con estas cuestiones. Por ejemplo, si un sensor de presión tiene un campo de medida de 0-1000 kPa, la salida está comprendida entre 0-1 V y la exactitud es del ±1% sobre el fondo de escala, la presión medida en cualquier punto puede presentar un error de ±10 kPa. También se suele expresar en % sobre el valor de la lectura. No obstante, una definición correcta se debería realizar en términos estadísticos. Si por ejemplo, el fabricante encontró que el error seguía una distribución normal, la desviación típica, σ indicaría el margen de error en el que está comprendido el 68% de las lecturas. El 95% de las mismas estaría comprendida entre $\pm 2\sigma$, y aunque es poco probable, es posible que alguna medida estuviera comprendida entre un margen mucho mayor. Por lo tanto, al indicarnos el fabricante la exactitud en % nos está diciendo el margen de error en el que están comprendidas la mayor parte de las lecturas (habitualmente, el 95%), pero no todas.

1.3.3 PROPAGACIÓN DE ERRORES

Los sistemas están compuestos por subsistemas que presentan errores. ¿Cómo cuantificar el error total? Supongamos que se tienen dos resistencias en serie, una de 1000 Ω ± 5% y otra de 100 Ω ± 10%. ¿Cuál es el valor de la resistencia total? En principio, el valor máximo de la resistencia equivalente será:

$$R_{eq-máx} = (1000 + 0,05 \cdot 1000) + (100 + 0,1 \cdot 100) = 1100 + 60$$
$$R_{eq-mín} = (1000 - 0,05 \cdot 1000) + (100 - 0,1 \cdot 100) = 1100 - 60$$

es decir, podríamos decir que la resistencia equivalente es de 1100 ± 60 Ω (1100 Ω ± 5,5%). Sin embargo, es muy improbable que ambas resistencias presenten sus valores máximos y mínimos simultáneamente (dependerá de la distribución estadística del error) por lo que se estaría sobrestimando el error. En general, cuando se suman dos números con error a ± k_1a y b ± k_2b, el resultado debería cuantificarse en términos estadísticos. Una aproximación al resultado, generalmente válida, teniendo en cuenta una distribución del error normal es: (a + b) ±e_r, donde el error, e_r, es [1.14]:

$$e_r = \sqrt{(k_1 a)^2 + (k_2 b)^2} \tag{1.6}$$

así, aplicando la fórmula al ejemplo, el resultado sería 1100 ± 51 Ω (1100 Ω ± 4,64%).

Para el resto de las operaciones (resta, producto y cociente), los resultados se muestran en la tabla 1.2. Observe que operaciones como la resta y el cociente producen una gran propagación del error relativo. Por ejemplo, la diferencia entre las dos resistencias consideradas en el presente apartado es 900 ± 51 Ω (900 Ω ± 5,67%), mayor que en el caso de la suma.

Operación		Resultado	Resultado corregido
Suma	$(a \pm k_1 a) + (b \pm k_2 b)$	$(a + b) \pm (k_1 a + k_2 b)$	$(a + b) \pm \sqrt{(k_1 a)^2 + (k_2 b)^2}$
Diferencia	$(a \pm k_1 a) - (b \pm k_2 b)$	$(a - b) \pm (k_1 a + k_2 b)$	$(a - b) \pm \sqrt{(k_1 a)^2 + (k_2 b)^2}$
Producto*	$(a \pm k_1 a) \cdot (b \pm k_2 b)$	$ab \pm ab(k_1 + k_2)$	$ab \pm ab\sqrt{k_1^2 + k_2^2}$
Cociente*	$\frac{a \pm k_1 a}{b \pm k_2 b}$	$\frac{a}{b} \pm \frac{a}{b}(k_1 + k_2)$	$\frac{a}{b} \pm \frac{a}{b}\sqrt{k_1^2 + k_2^2}$

Tabla 1.2. Resultados de las operaciones básicas de números con error (adaptado de [1.14]).
* (Resultados cuando los errores porcentuales son pequeños)

EJERCICIO 1.3

Se pretende determinar la potencia consumida por un equipo midiendo la corriente y la tensión. El amperímetro presenta un campo de medida de 0-50 A y un error sobre la lectura de un ±2%. El voltímetro presenta un fondo campo de medida de 0-100 V y un error sobre la lectura del ±1%. Al efectuar la medición, el polímetro indica 15 A y el voltímetro 60 V, por lo que la potencia consumida es, en principio, 900 W. ¿Cuál es el error?

SOLUCIÓN:

De la tabla 1.2 se obtiene fácilmente, la potencia de salida es de 900 W ± 3% cuando se emplea la expresión sin corregir o 900 W ± 2,24% cuando se emplea el resultado corregido.

1.3.4 CALIBRACIÓN

La calibración de un sistema consiste en establecer, con la mayor exactitud posible, la correspondencia entre las indicaciones de un instrumento de medida y los valores de la magnitud que se mide con él. Las tolerancias de los componentes y las no idealidades de los circuitos electrónicos conducen a que dos sensores o dos instrumentos de medida aparentemente idénticos no presenten nunca curvas idénticas de calibración. Además, la curva de calibración de los instrumentos varía a lo largo del tiempo y del grado de utilización de los mismos.

En los sistemas con curvas de calibración lineales, las diferencias más notables se suelen centrar en dos aspectos: el nivel (*offset*) y la sensibilidad (figura 1.26). Estas diferencias conducen a errores sistemáticos que pueden ser minimizados siempre que exista posibilidad de calibrar el equipo.

Los métodos de calibración más sencillos son la calibración a un punto y la calibración del cero y de la sensibilidad:

- **Calibración a un punto:** consiste en actuar sobre el sistema de medida de forma que para un punto concreto la salida sea lo más exacta posible. Por ejemplo, las básculas de baño analógicas incorporan un mando para ajustar el cero. Cuando el usuario observa que en ausencia de peso la indicación de la báscula es distinta de cero, actúa sobre el mando hasta que la medida sea correcta. Las básculas digitales no incorporan aparentemente dicha posibilidad; en realidad lo hacen automáticamente, es decir, memorizan el resultado en ausencia de peso, aunque no se muestra al usuario, y lo restan a todas las medidas.

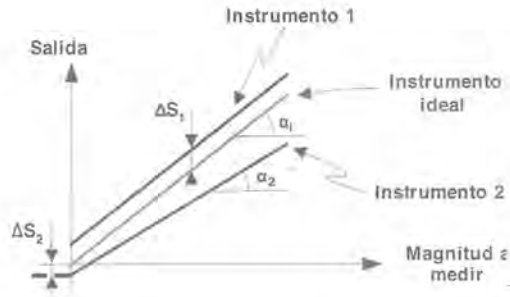


Figura 1.26. Representación de la curva de calibración ideal y la que corresponde a dos instrumentos reales. Se puede observar que el instrumento 1 proporciona una salida que siempre es un cierto nivel ΔS_1, mayor que el ideal. El instrumento 2, además de presentar un cierto nivel menor ΔS_2, tiene una sensibilidad —$\tan(\alpha_2)$— diferente a la ideal —$\tan(\alpha_i)$—.

En muchos sistemas, la calibración a un punto suele efectuarse sobre el valor cero de la variable de entrada porque suele ser uno de los puntos para los que más fácilmente suele ser conocido el valor verdadero, como en el caso del ejemplo de la báscula, los velocímetros, acelerómetros, sistemas para medir desplazamiento, etc.

- **Calibración del cero y de la sensibilidad:** para ajustar perfectamente una curva de calibración lineal se necesitarían ajustar dos puntos o un punto y la pendiente o sensibilidad. Muchos instrumentos de medida incorporan esta posibilidad de calibración aunque no siempre sea accesible por el usuario mediante mandos externos, aunque sí por el fabricante. El ajuste se realiza mediante dos mandos, uno permite modificar el nivel (*offset*) y el otro la sensibilidad o ganancia del sistema (*gain*). El procedimiento debe realizarse ajustando primero el cero (figura 1.27b) y, a continuación, la ganancia (figura

1.27c). Para este último ajuste es necesario medir en otro punto y ajustar la ganancia de forma que en este segundo punto la salida sea la deseada.

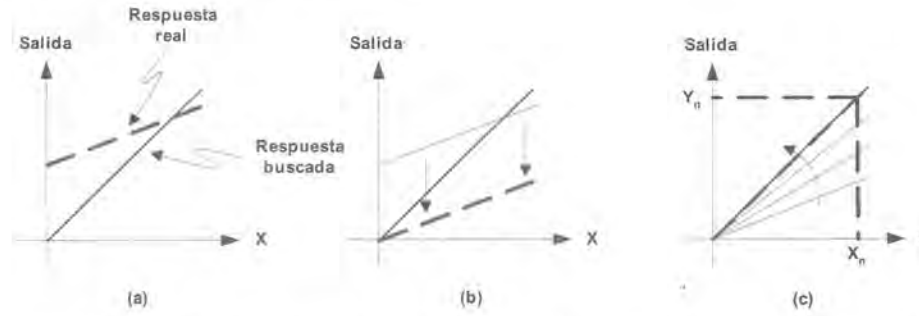


Figura 1.27. Calibración del cero y de la ganancia. (a) Situación original. (b) En el primer paso se ajusta el nivel de cero. (c) Seguidamente se mide a otro punto y se ajusta la ganancia hasta que la salida coincida con la esperada.

EJERCICIO 1.4

Se dispone de una báscula industrial para medir entre 0 y 1000 kg. Después de un uso repetido de la misma se han observado desviaciones en la medida (por ejemplo, en ausencia de peso indica 10 kg) por lo que se ha decidido recalibrar el sistema. La báscula permite ajustar el cero y la ganancia.

¿Sería correcto ajustar en primer lugar la ganancia y después el cero?

SOLUCIÓN:

No.

Si, por ejemplo, la situación original es la mostrada en la figura 1.28a y se ajusta en primer lugar la salida en un punto de medida X_n mediante la modificación de la ganancia (figura 1.28b) y, a continuación, se ajusta el nivel en el cero, la curva de calibración resultante (figura 1.28c) presenta un error considerable.

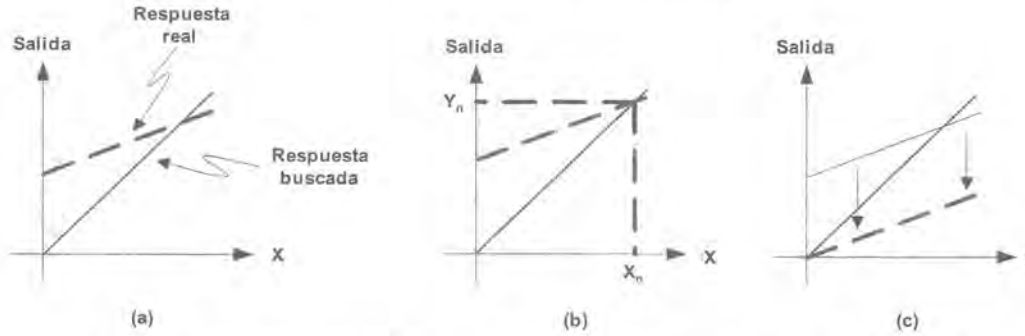


Figura 1.28. (a) Situación original. (b) Resultado de haber aumentado la ganancia hasta que la salida en el punto X_n coincida con la deseada. (c) Curva de calibración resultante después de ajustar el cero en el siguiente paso.

La utilización de microcontroladores y microprocesadores permite técnicas de calibración más complejas como la calibración a varios puntos realizando el ajuste mediante mínimos cuadrados o mediante el autoajuste del equipo realizando una calibración a dos puntos cualesquiera.

No obstante, existe un problema inherente a todo proceso de calibración consistente en que para calibrar un equipo es necesario conocer el valor asumido como exacto de la variable de entrada. Si bien es cierto que algunos puntos son en ocasiones fáciles de determinar (como, por ejemplo, el cero en una báscula) generalmente un usuario final no dispone de los medios necesarios para calibrar otros puntos. Para calibrar un equipo se necesita otro que presente una exactitud mayor que el equipo que se pretende calibrar o bien, recurrir a patrones o ensayos de referencia (por ejemplo, si se desea calibrar el cero de un termómetro, se podría determinar el valor exacto a través del punto de congelación del agua). Por ello, la calibración adecuada de un equipo sólo suele estar al alcance de grandes empresas que disponen de laboratorios específicos o de determinados organismos como el Centro Español de Metrología[4].

1.4 Características dinámicas

1.4.1 FUNCIÓN DE TRANSFERENCIA

Cualquier sistema físico puede ser considerado como un bloque con entradas, salidas y una expresión que las relaciona. En una primera aproximación, esta expresión más o menos compleja se considera como independiente del tiempo, determinando la caracterización estática de dicho sistema como se comentó en el apartado anterior; sin embargo, en condiciones más reales, el principio de causalidad obliga a que las salidas vayan siempre detrás de las entradas, es decir, las salidas serán una respuesta ante las excitaciones de entrada.

A este principio básico no escapan los sistemas de instrumentación electrónica, considerados como bloques con una o más variables de entrada que son excitaciones y una o más variables de salida que son respuestas —no inmediatas— ante esas excitaciones. Para simplificar, consideremos el caso de un sistema cualquiera con una variable de entrada y una variable de salida como el de la figura 1.29a.

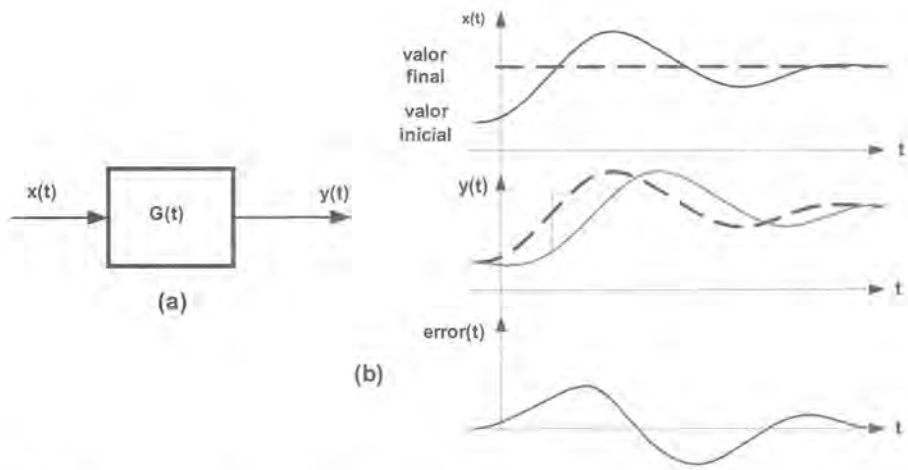


Figura 1.29. (a) Concepto general de función de transferencia y, (b) señales de entrada y salida teniendo en cuenta el "retraso y modificación" impuesto por el sistema.

[4] El CEM fue creado por Real Decreto 415/1985, de 27 de marzo y está situado en la localidad de Tres Cantos (Madrid). Algunas de sus funciones son las siguientes:
- Custodia y conservación de los patrones nacionales de las unidades de medida.
- Establecimiento de las cadenas oficiales de calibración.
- Ejercicio de las funciones de la Administración General del Estado en materia de metrología legal.
- Ejecución de proyectos de investigación y desarrollo en el ámbito metrológico.
- Formación de especialistas en Metrología.

El efecto de la no inmediatez en la respuesta del sistema se traduce en una diferencia entre el valor esperado en cada momento y el que realmente se produce, lo que se podría modelar como un "error momentáneo" del sistema de medida que se manifiesta cuando hay cambios en la variable de entrada (figura 1.29b). Entonces, en sistemas que presenten una entrada que esté variando constantemente, la salida lo hará también pero con un retraso (figura 1.30a) impuesto por la causalidad del sistema, mientras que en otros sistemas en que la entrada varíe en un momento dado (pasa de un valor a otro), pasado un tiempo se logrará alcanzar el valor final que corresponda (figura 1.30b).

Teniendo en cuenta que el objetivo de un sistema de instrumentación es aproximarse todo lo que sea posible a la obtención del valor real que se pretende medir, estos efectos introducen una desviación momentánea o permanente que afecta a la medida. La pregunta que nos podemos hacer es: ¿resulta importante este efecto?

La respuesta no es general, sino que depende de cada caso: en referencia a las señales de la figura 1.30a, si la información es la amplitud o la frecuencia, el efecto del retraso carecerá de importancia pero, si la información es la fase o el valor instantáneo, se estará introduciendo un error sistemático que deberá ser tenido en cuenta. En el caso de la figura 1.30b, la señal puede considerarse correcta durante la mayor parte del tiempo, excepto en el intervalo de cambio, y la importancia del efecto se traduce en un tiempo de espera hasta obtener el valor "real" después de cada cambio.

La ciencia matemática proporciona algunas herramientas de análisis útiles basadas en el uso de las transformadas de Laplace o de Fourier [1.12] que permiten modelar el comportamiento de un sistema lineal mediante una función de la frecuencia compleja s o de la frecuencia jω denominada función de transferencia. Así, la salida del sistema y(t) se puede obtener mediante la función de transferencia del sistema G(s) o G(jω) que se aplica a la variable de entrada x(t):

$$y(j\omega) = G(j\omega)\cdot x(j\omega) \qquad y(s) = G(s)\cdot x(s) \tag{1.7}$$

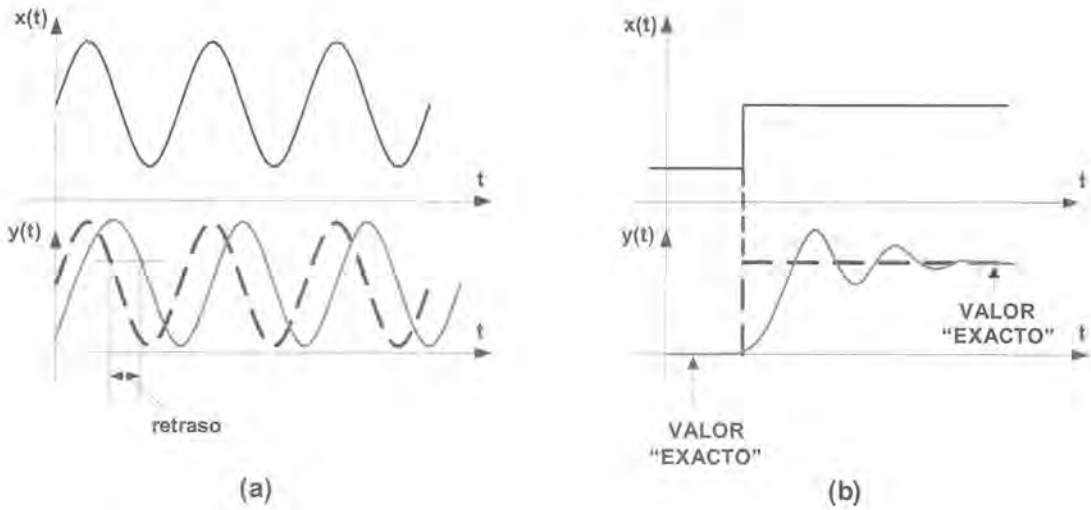


Figura 1.30. Efectos sobre la señal de salida de un sistema: (a) efecto de retraso permanente sobre una señal senoidal; (b) efecto transitorio sobre un cambio de nivel.

Las funciones de transferencia G(s) y G(jω) son expresiones racionales iguales (una en s y otra en jω) con un polinomio en el numerador y otro en el denominador; este último será de mayor o igual grado que el

primero en sistemas causales como cualquiera de los que nos podemos encontrar en el ámbito de la ingeniería [1.12], [1.15].

Estas funciones pueden llegar a resultar muy complicadas, presentando polinomios de grado muy alto, si se quieren recoger todos los efectos propios del sistema. En la práctica, la mayoría de los sistemas quedan bien representados por funciones de transferencia cuyo grado máximo sea dos o tres.

La función de transferencia dependiente de la frecuencia jω representa el comportamiento del sistema para cualquier frecuencia de la señal de entrada y es una forma muy común de caracterizar el comportamiento de los sistemas electrónicos. Resulta una expresión en el plano complejo con una parte real y una imaginaria, aunque se suele preferir usar la representación mediante el módulo y la fase. Consideremos, por ejemplo, la siguiente función de transferencia correspondiente a un determinado sistema:

$$|G(j\omega)| = \frac{\omega^2 + 3\omega + 5}{2\omega^3 + \omega^2 + 6\omega + 10}$$

La función recoge el comportamiento del sistema a todas las frecuencias; así, si en la anterior expresión hacemos ω = 0 rad/s lo que significa una entrada de continua, la función nos indica que el módulo de la ganancia del sistema es 1/2 mientras que a la frecuencia ω = 100 rad/s, el módulo de la ganancia del sistema resulta un poco mayor de 0,005. Lo mismo se podría hacer para todas las demás frecuencias, pudiendo representarse en un diagrama en que se coloca en el eje horizontal la frecuencia y, en el vertical, el módulo de la ganancia.

EJERCICIO 1.5

Sea un sistema definido por el módulo de su función de transferencia $G(\omega) = (\omega^2 + 30\omega + 50)/(\omega^3 + \omega + 10)$. *Represente de forma aproximada el módulo de la ganancia en decibelios en función del logaritmo de la frecuencia ω.*

SOLUCIÓN:

Como la frecuencia se va a representar en un eje logarítmico, vamos dando valores que se vayan espaciando más o menos igual a lo largo del eje: ω = 0,01, 0,02, 0,04, 0,08, 1, 2, 4, 8, y así, sucesivamente. Hay que notar que el valor de continua ω = 0 no puede ser representado en el eje horizontal. En la figura 1.31 se muestra una tabla de valores y la representación correspondiente.

Log ω	|G(ω)|
0,01	5,0001
0,02	5,0001
0,04	5,0004
0,08	5,0028
0,1	5,0077
0,2	5,0129
0,4	5,0359
0,8	5,2759
1	5,7454
2	6,1716
4	6,2123
8	0,8603
10	0,3072
20	0,1890
40	0,0769
80	0,0185
100	0,0104

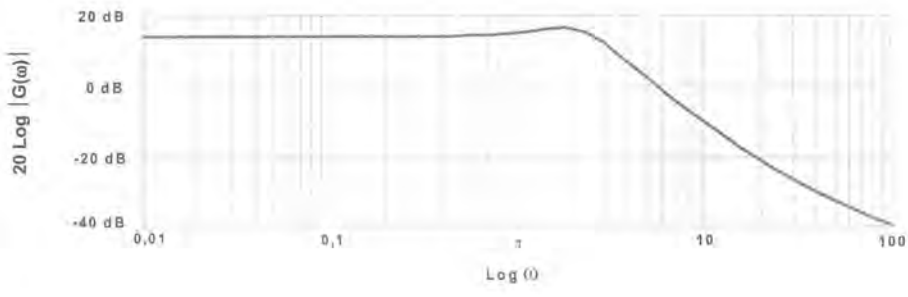


Figura 1.31. Tabla de valores y representación de la función del ejercicio 1.5.

1.4.2 CARACTERIZACIÓN DE LA FUNCIÓN DE TRANSFERENCIA

La obtención de la función de transferencia de cualquier sistema resulta una tarea muy interesante puesto que permite caracterizar su comportamiento dinámico; es lo que se conoce como modelado dinámico. Para modelar un sistema existen varias posibilidades aunque ninguna de ellas resulte simple en la mayoría de los casos:

a) Modelado teórico: consiste en extraer las relaciones teóricas entre las variables del sistema, su linealización en torno a un punto de funcionamiento y la aplicación de la transformación de Fourier (o de Laplace) para conseguir su función de transferencia. Suele ser una labor tediosa en muchos casos y la necesidad de realizar aproximaciones y linealizaciones le impide alcanzar gran exactitud.

b) Modelado empírico: consiste en someter el sistema a determinadas excitaciones en la entrada y observar su salida, que muestra una buena parte del comportamiento del sistema. Esta técnica se puede llevar a cabo de varias formas diferentes: introduciendo un barrido de todas las frecuencias en la señal de entrada con lo que obtenemos punto a punto, y para cada una de esas frecuencias, la relación salida/entrada, introduciendo ruido blanco[5] u observando la respuesta del sistema ante una entrada impulsional o en escalón [1.12], [1.15] según se muestra en la figura 1.32.

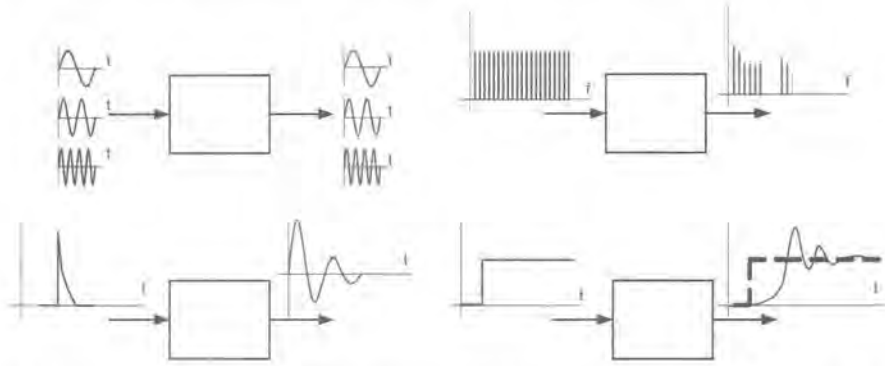


Figura 1.32. Obtención empírica de la respuesta de un sistema para caracterizar su función de transferencia mediante diversas técnicas.

Una vez obtenida la función de transferencia del sistema con alguno de los anteriores métodos o con una combinación de ambos, tenemos la representación matemática del comportamiento dinámico del sistema; la pregunta que viene ahora es obvia: ¿cómo utilizamos esa función para definir de una manera más cómoda nuestro sistema?

Para contestar a esta cuestión, introduciremos otro concepto que es la representación de la función de transferencia en función de la frecuencia ω, representación que se denomina diagrama de Bode y que dibuja el módulo (en dB) y la fase (en °) de $G(j\omega)$ en función del logaritmo de la frecuencia. Representaciones habituales en sistemas electrónicos se muestran en la figura 1.33.

[5] El ruido blanco constituye un ruido totalmente aleatorio (sin estructura ni información alguna) y se caracteriza porque su espectro frecuencial presenta una distribución plana lo que significa que todas las frecuencias tienen la misma participación.

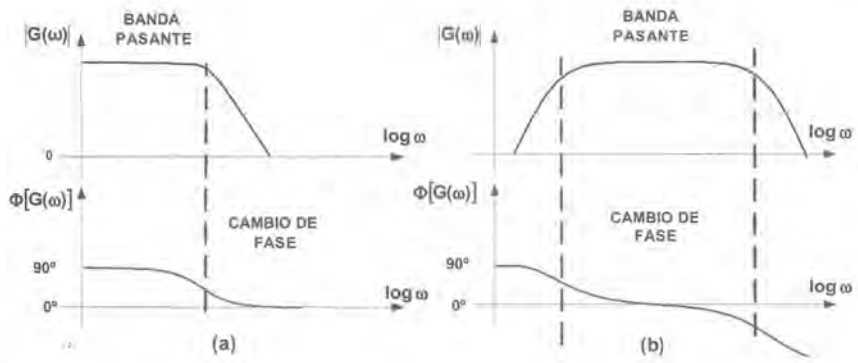


Figura 1.33. Representación del diagrama de Bode en dos casos muy habituales dentro de los sistemas electrónicos.

La interpretación que se puede hacer es la siguiente: todas las señales senoidales de frecuencia comprendida en la zona plana son amplificadas por igual por el sistema, mientras que las demás son amplificadas menos o atenuadas en un nivel mayor cuanto más se alejen de la zona plana.

El desarrollo en serie de Fourier permite descomponer toda señal periódica de periodo T en suma de senoides de periodos submúltiplos de T; cada una de esas componentes se puede representar por una línea vertical de valor igual a su amplitud situada en el punto correspondiente a su frecuencia (espectro de la señal). Teniendo en cuenta que el sistema es lineal, se puede aplicar superposición para concluir que, a la salida, se obtendrá una suma de todas las senoides que constituyen el espectro, amplificadas o atenuadas en función de su frecuencia. Así se obtiene la salida del sistema para esa señal de entrada (figura 1.34).

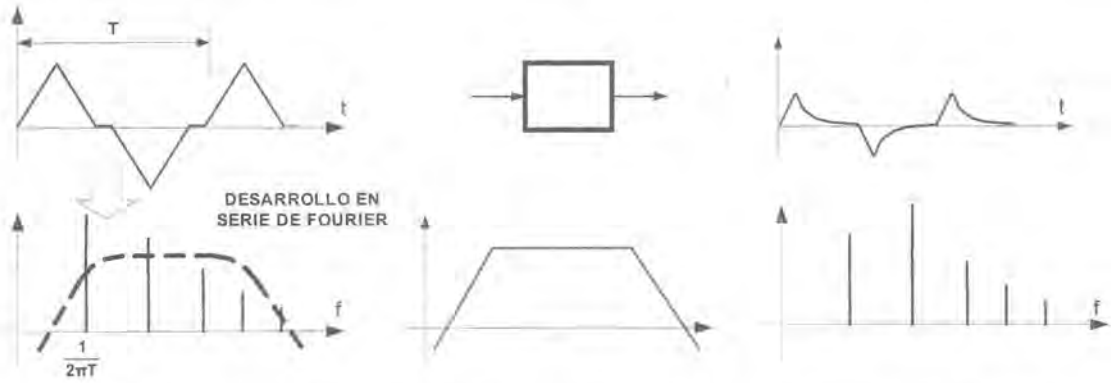


Figura 1.34. Utilización del diagrama Bode y el desarrollo en serie de Fourier para determinar la respuesta de un sistema ante una entrada cualquiera.

El resultado obtenido se traduce en una serie de cambios en el aspecto de la señal de salida respecto al aspecto de la señal de entrada a través de la diversa amplificación/atenuación y desfase (cambio respecto a la fase inicial de la señal introducido por el sistema) que sufre cada una de las componentes senoidales de la entrada. Esto se puede cuantificar en función de varios parámetros que permiten caracterizar el comportamiento del sistema:

a) La **distorsión de amplitud** es el efecto que se produce por el cambio en el espectro de la función de salida respecto al espectro de la entrada debido a cambios en la amplificación de cada una de las componentes.

b) La adición de otras señales no presentes en la entrada que se originan por un comportamiento no lineal del sistema. Este efecto se suele cuantificar mediante un parámetro globalizador como es la **distorsión armónica total** (*Total Harmonic Distorsion* o THD). De acuerdo con [1.10], [1.13] la THD del sistema se define como el cociente entre la suma de la potencia de todos los armónicos de frecuencias superiores a la frecuencia fundamental y la potencia del fundamental. Se suele expresar en dB o en porcentaje, siendo esta última la forma más habitual en el campo de la instrumentación electrónica y de la electrónica de potencia mientras que la primera es mucho más frecuente en el campo de los procesos de telecomunicación. Sobre una misma impedancia, se puede considerar sólo la corriente o sólo la tensión y la expresión de la THD resulta ser:

$$\mathrm{THD}(\%) = \frac{\sqrt{V_1^2 + V_2^2 + ..., V_n^2}}{V_f} 100 \qquad (1.8)$$

donde cada V_i representa el valor eficaz de cada uno de los armónicos obtenidos a la salida y V_f es el valor eficaz del fundamental.

c) La **distorsión de fase** es el cambio introducido en la fase de cada una de las componentes del espectro de entrada cuando atraviesan el sistema, como se muestra en la figura 1.33.

La importancia de estos efectos depende del caso, aunque los dos primeros suelen tener más trascendencia que el último. También puede ocurrir que cualquiera de estos efectos sean expresamente buscados, como en el caso de algunos sistemas como los de filtrado (véase la lección correspondiente a filtrado) en que el objetivo es precisamente la eliminación de determinadas componentes no deseadas.

Considerando sólo el comportamiento lineal de los sistemas (no hay distorsión armónica) y simplificando en consecuencia el anterior análisis, se puede llegar a dividir la respuesta del sistema en dos zonas: la que corresponde a la zona plana en que todas las señales son amplificadas en magnitud en igual cuantía y las zonas en que se produce una amplificación menor. Al establecer esa separación en valor de 3 dB (la potencia de la salida se reduce a la mitad respecto a la potencia que se obtendría en la zona plana), se puede decir que la zona en que la amplificación no varía más de 3 dB es la banda pasante del sistema y está limitada por una o dos frecuencias de corte (frecuencias a las que la amplificación es 3 dB menos que la que corresponde a la banda pasante). En la figura 1.35 se muestran las características de ganancia y de fase de dos sistemas diferentes, correspondientes al caso en que tenga una frecuencia de corte superior, ω_H (a) y al caso en que tenga una frecuencia de corte superior y otra inferior, ω_L (b). En el primer caso, el sistema amplifica hasta las señales de continua ($\omega = 0$) y el ancho de banda es $B = \omega_H$, mientras que en el segundo sólo amplifica señales entre las dos frecuencias de corte y $B = \omega_H - \omega_L$.

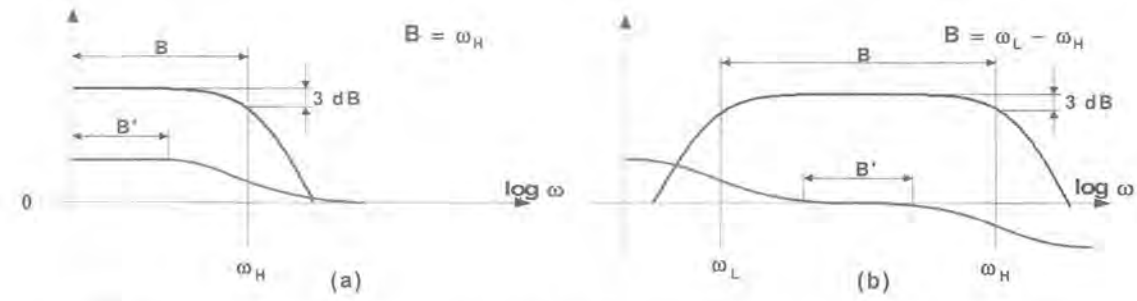


Figura 1.35. Concepto de ancho de banda sobre los diagramas de Bode de dos sistemas.

Como se puede observar en la figura 1.35, la fase cambia incluso dentro del ancho de banda del sistema; si no fuese deseable un cambio de fase, la zona útil de trabajo del sistema resultaría menor que el ancho de banda. Aparece así el concepto de **ancho de banda útil** del sistema B' y que puede coincidir con B (no importa el desfase) o ser menor (sí importa el desfase).

EJERCICIO 1.6

Determinar el ancho de banda del circuito de la figura 1.36. ¿Qué salida se obtendría con una entrada de tensión $v_i(t) = 10 \text{ sen } (10\ t) + 10 \text{ sen } (1000\ t)$?

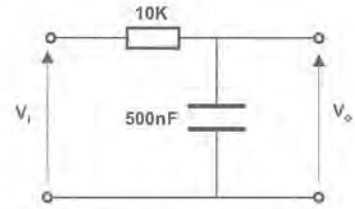


Figura 1.36. Circuito para el ejercicio 1.6.

SOLUCIÓN:

El sistema tiene una función de transferencia $v_o/v_i = 1/(1 + j0{,}05\omega)$ cuya representación gráfica aparece en la figura 1.37. El ancho de banda es equivalente a la frecuencia de corte superior y se calcula para el punto en el que la ganancia es de –3 dB y que resulta ser B = 20 rad/s.

La señal de entrada posee dos componentes, de frecuencias 10 y 1000 rad/s; la primera atraviesa el sistema con ganancia unidad, mientras que la segunda es atenuada en un valor de 0,02, es decir, unos 34 dB, ya que cae fuera del ancho de banda del circuito.

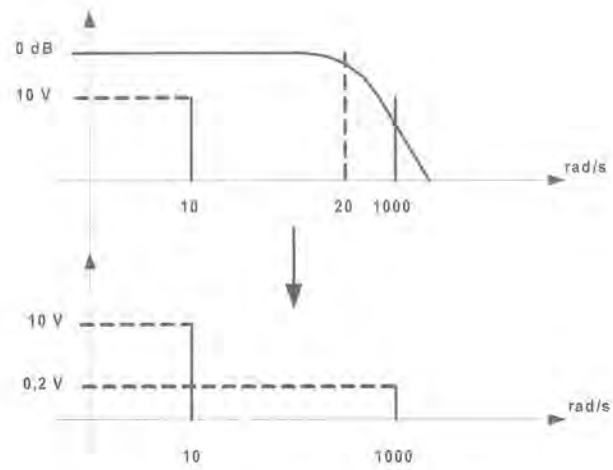


Figura 1.37. Diagrama de Bode del circuito de la figura 1.36 y su efecto sobre la señal de entrada.

En cuanto a la fase que presentarán las señales en la salida, la primera (10 rad/s) sufrirá un desfase de $-30^\circ = -\pi/6$ rad, mientras que la segunda (1000 rad/s), que está totalmente fuera del ancho de banda, sufrirá el máximo desfase, es decir, $-90^\circ = -\pi/2$ rad.

Por lo tanto, la salida que se obtendría sería $v_o(t) = 10 \text{ sen } (10\ t - \pi/6) + 0{,}2 \text{ sen } (1000\ t - \pi/2)$.

1.4.3 EVALUACIÓN DE LA RESPUESTA DINÁMICA

En el apartado anterior hemos caracterizado la función de transferencia en el dominio de la frecuencia, observando los cambios que introducía en las variables que atravesaban un determinado sistema para generar su respuesta en todas las frecuencias.

Si bien la respuesta dinámica de un sistema puede establecerse desde el punto de vista de los conceptos antes estudiados, también son posibles otras caracterizaciones en función de conceptos que se refieran directamente al tiempo en lugar de referirse a la frecuencia. En este sentido, tiene especial importancia en el ámbito de la instrumentación, la respuesta que proporciona un determinado sistema o equipo en su salida ante un cambio brusco en la variable de entrada (escalón de entrada) ya que incorpora todos los efectos dinámicos propios del sistema. La rapidez en el cambio de la salida ante un cambio en la entrada y el aspecto que presenta durante el cambio la variable de salida permiten evaluar el comportamiento del sistema de la misma forma que se hacía con el ancho de banda.

La respuesta de un sistema ante un cambio en la entrada se traduce en un cambio en la salida con un cierto retraso y una determinada evolución desde el valor anterior hasta el nuevo (figura 1.38).

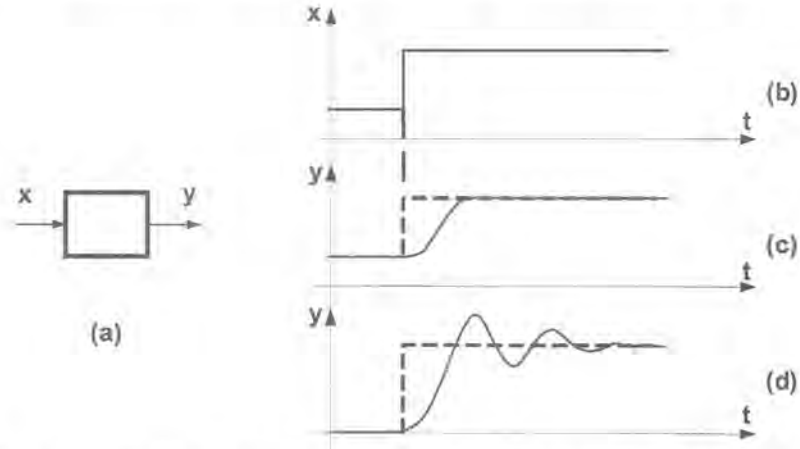


Figura 1.38. Respuestas de un sistema (a) ante una entrada en escalón (b), (c) sistema lento y, (d) sistema rápido y poco amortiguado.

La respuesta de un sistema ideal que no incorporase dinámica alguna correspondería a la respuesta punteada mientras que la de un sistema real puede tener un aspecto similar a la de la figura 1.38c o 1.38d, en las que se ve cómo la salida alcanza finalmente el mismo valor previsto en el caso ideal (permanente) aunque pasado un determinado tiempo (transitorio). Aunque los sistemas pueden tener muchos tipos de respuestas ante un escalón de entrada y que están directamente relacionadas con el orden del numerador y denominador de su función de transferencia lo cierto es que, en la inmensa mayoría de los casos, la respuesta es similar a la que presentaría un sistema de primer orden en el denominador o de segundo orden en el denominador y que corresponden de forma aproximada a las representadas en la figura 1.38c y 1.38d. Así pues, vamos a tratar de caracterizar cómo se comportan estos sistemas en cuanto a su respuesta ante entrada en escalón.

Sistemas de primer orden

La respuesta de un sistema de primer orden ante una entrada en escalón corresponde con una función temporal como ésta:

$$x(t) = X_F\left(1 - e^{-\frac{t}{\tau}}\right) \tag{1.9}$$

y se representa en la figura 1.39. El valor de X_F corresponde a su valor final (el que se alcanzaría hipotéticamente en el infinito) y τ es el valor de la constante de tiempo. Esta constante representa el tiempo que tarda la salida en alcanzar el 63% de su valor final.

Por tanto, el parámetro dinámico que define un sistema de primer orden es la **constante de tiempo** aunque se pueden definir otros parámetros que también permiten caracterizar lo rápido que resulta un sistema de primer orden, tales como:

a) El **tiempo de subida** (*rise time*), t_r definido como el tiempo que transcurre entre que el sistema alcanza el 10% y el 90% del valor final.

b) El **tiempo de establecimiento** (*settling time*), t_s definido como el tiempo que transcurre hasta que el sistema proporciona una salida dentro del margen de tolerancia definido por su precisión[6].

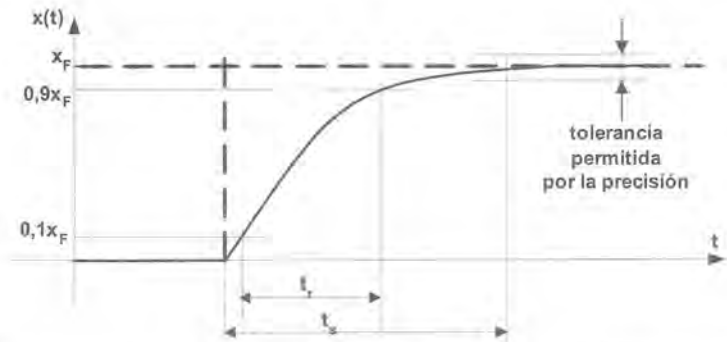


Figura 1.39. Respuesta de un sistema de primer orden ante una entrada en escalón.

Sistemas de segundo orden

En los sistemas de segundo orden, la respuesta ante una entrada en escalón no tiene un aspecto único, sino que pueden presentarse tres casos diferentes según la inercia y la amortiguación que presente el sistema. Tenemos:

a) *Sistemas sobreamortiguados*, sistemas lentos cuya respuesta es similar a la de un sistema de primer orden (figura 1.40a).

b) *Sistemas subamortiguados*, sistemas rápidos que presentan oscilaciones por encima del valor final (figura 1.40b).

c) *Sistemas con amortiguamiento crítico*, que están en el límite entre los dos comportamientos anteriores, más rápidos que los primeros pero con un aspecto de respuesta muy parecido (figura 1.40c).

[6] El tiempo de establecimiento se puede definir de otras formas; por ejemplo, es frecuente el uso de cinco veces la constante de tiempo para determinarlo, sobre todo en el ámbito de la regulación automática. Aquí hemos preferido utilizar una definición instrumental [1.12].

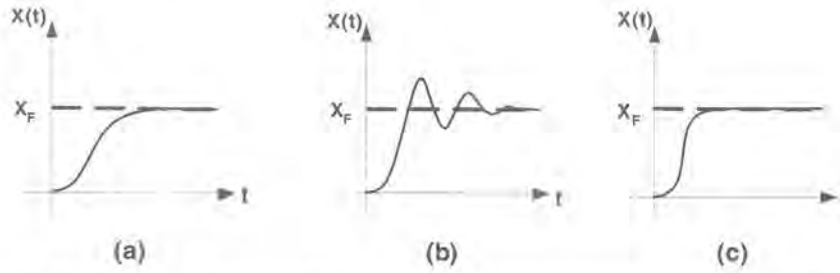


Figura 1.40. Sistemas de segundo orden: (a) sobreamortiguado, (b) subamortiguado y, (c) con amortiguamiento crítico.

La definición de la respuesta de un sistema de segundo orden es, por tanto, dependiente de la forma real que presente lo que significa que, en términos generales, se hablará de uno u otro parámetro en función de ella. No obstante, cuando la respuesta corresponde a los tipos sobreamortiguado o con amortiguamiento crítico los parámetros a utilizar pueden ser los mismos que en los sistemas de primer orden, tiempo de subida y tiempo de establecimiento ya que la respuesta sólo se diferencia significativamente de aquel caso en la pendiente en el origen que ahora es nula.

Cuando estemos en el caso de sistemas poco amortiguados (figura 1.40b), la presencia de sobreoscilaciones (momentos en los que se sobrepasa el valor final) hace que el término tiempo de subida no suela tener mucho sentido mientras que sí que puede ser importante conocer cuánto se llega a sobrepasar el valor final. Para este caso se definen:

a) **Sobreoscilación** (S_u), definida como el cociente entre el valor máximo que se alcanza en el tiempo t_u (figura 1.41) y el valor final en tanto por uno o en porcentaje.

$$S_u(\%) = \frac{X_{máx}}{X_F} 100 \tag{1.9}$$

No suelen interesar valores altos de sobreoscilación ya que puede poner en peligro los sistemas por valores excesivos no tolerables de la señal.

b) **Tiempo de establecimiento** (t_s), parámetro similar al definido para los sistemas de primer orden pero, en este caso, como quiera que la señal se acerca y se aleja varias veces del valor final, se define como el tiempo a partir del cual la señal queda siempre dentro de la banda del margen de confianza definida por la precisión del sistema [1.17].

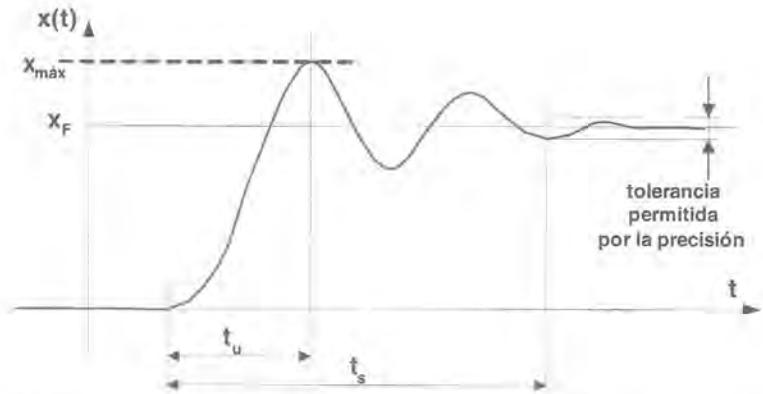


Figura 1.41. Respuesta de un sistema de segundo orden ante una entrada en escalón.

Hay que tener en cuenta que si bien la respuesta temporal de un sistema lo define unívocamente y contempla aspectos tanto lineales como no lineales, la asimilación de esa respuesta a un sistema de primer o de segundo orden es sólo una aproximación de la que se pueden extraer algunos parámetros característicos como los que hemos comentado. Así pues, usando sólo esos datos no definimos el sistema pero, al menos, extraemos de él aquello que nos resulta más útil para decir cómo se comporta desde el punto de vista dinámico, es decir, lo rápido que es y si presenta o no sobreoscilaciones.

Existe una relación entre el comportamiento frecuencial y temporal de cualquier sistema que indica que sistemas con frecuencias de corte superiores elevadas resultan rápidos y tienen bajos tiempos de subida y establecimiento; por el contrario, cuando las frecuencias de corte superior son reducidas, los sistemas resultan lentos y los tiempos de subida y de establecimiento son largos. Los sistemas que presentan sobreoscilaciones pronunciadas suelen tener picos de resonancia en sus diagramas de Bode por lo que también puede hacerse un traspaso de ideas entre el dominio de la frecuencia y del tiempo.

Como criterio general, se prefiere que los equipos de medida tengan una buena velocidad de respuesta con tiempos de subida y establecimiento cortos y, sobre todo, cuando están inmersos en un sistema de control como sensor de una determinada variable; en este caso, la dinámica del equipo de medida puede provocar serios problemas en el funcionamiento del proceso en bucle cerrado y dificultar la tarea del regulador [1.12].

EJERCICIO 1.7

Tenemos un termómetro de mercurio que está constituido por un tubo de vidrio de conductividad térmica k y que presenta un espesor de pared d, superficie total S y masa térmica despreciable en cuyo interior hay una masa de mercurio m con calor específico c_e [1.7]. Determine la función de transferencia del sistema, así como la constante de tiempo y el tiempo de subida de la respuesta cuando se produce un escalón térmico en la entrada desde 20ºC hasta 50ºC.

SOLUCIÓN:

Partimos de una situación de equilibrio en la que tanto el ambiente exterior como el mercurio del interior están a 20ºC. En estas condiciones la temperatura exterior cambia bruscamente a 50ºC (por ejemplo, sumergimos el termómetro en un líquido a esa temperatura o, simplemente, lo cambiamos de sitio).

El flujo de calor q(t) que atraviesa la superficie de vidrio es proporcional a la diferencia de temperaturas y viene dado por la expresión:

$$q(t) = \frac{kS}{d}\left(T_A(t) - T_{Hg}(t)\right)$$

donde $T_A(t)$ es la variable que representa la temperatura ambiente y $T_{Hg}(t)$ representa a la temperatura interior del mercurio.

Suponiendo que la masa térmica del vidrio es despreciable, todo ese flujo de calor es absorbido por la masa térmica del mercurio siguiendo la expresión:

$$q(t) = mc_e \frac{dT_{Hg}(t)}{dt}$$

Sustituyendo el valor de q(t) en una de ellas, se obtiene la expresión:

$$mc_e \frac{dT_{Hg}(t)}{dt} = \frac{kS}{d}\left(T_A(t) - T_{Hg}(t)\right)$$

con lo que podemos encontrar la influencia de la entrada, $T_A(t)$ sobre la salida $T_{Hg}(t)$ según la siguiente ecuación diferencial:

$$\frac{dT_{Hg}(t)}{dt}+\frac{kS}{dmc_e}T_{Hg}(t)=\frac{kS}{dmc_e}T_A(t)$$

La resolución de una ecuación diferencial en variables separadas como esta es muy sencilla y ya se puede ver que el comportamiento va a corresponder a una subida exponencial como la de la expresión (1.9). No obstante, proseguiremos para buscar el modelo matemático del sensor:

Si al término kS/dmc_e se le denomina A, lo que es un parámetro que depende sólo del sistema, y se aplica la transformada de Laplace, se obtiene la expresión:

$$sT_{Hg}(s)+AT_{Hg}(s)=AT_A(s)$$

o, lo que es lo mismo, el sistema queda caracterizado por la siguiente función de transferencia en el dominio s:

$$G(s)=\frac{T_{Hg}(s)}{T_A(s)}=\frac{A}{A+s}=\frac{1}{1+s/A}$$

Si se trabajase en el dominio de la frecuencia, la función de transferencia sería la misma pero con $j\omega$ en lugar de la variable s, es decir:

$$G(j\omega)=\frac{T_{Hg}(j\omega)}{T_A(j\omega)}=\frac{A}{A+j\omega}=\frac{1}{1+j\omega/A}$$

La constante de tiempo para el sistema resulta ser 1/A lo que implica que la respuesta temporal del sistema cuando pasa de 20°C a 50°C seguirá la expresión[7]:

$$T_{Hg}(t)=(50^\circ-20^\circ)\left(1-e^{-\frac{t}{A}}\right)+20^\circ$$

El valor final, tras el escalón de 30°C, será 50°C lo que ocurriría para $t=\infty$ pero el tiempo de subida se calculará resolviendo la expresión para $T_{Hg}=30^\circ\cdot0{,}9+20^\circ$ (tiempo para el 90% del valor final) y $T_{Hg}=30^\circ\cdot0{,}1+20^\circ$ (tiempo para el 10% del valor final):

$$t_{90\%}=2{,}30A \quad t_{10\%}=0{,}105A$$

Lo que fija un tiempo de subida de 2,19 veces la constante de tiempo del sistema. Podría calcularse también el tiempo de establecimiento siempre que se nos facilitara el dato de la tolerancia que se admite para el valor de la salida en función de la precisión del sistema.

EJERCICIO 1.8

En el ejercicio anterior, calcule el tiempo de establecimiento considerando que la tolerancia permitida para la lectura es del 2%. ¿Cómo se incrementaría el tiempo de establecimiento si la tolerancia permitida fuese sólo del 0,1%?

[7] Todos los sistemas de primer orden tienen una respuesta ante una entrada en escalón que se puede estimar siguiendo la expresión:

$$Y(t)=(Y_F-Y_O)\left(1-e^{-\frac{t}{\tau}}\right)+Y_O$$

donde Y_F es el valor final (en régimen permanente), Y_O es el valor inicial y τ, la constante de tiempo de la exponencial.

SOLUCIÓN:

A partir de la expresión de la respuesta del sistema, el 2% del valor final que corresponde al escalón de subida se produce a un nivel de $50 - 0{,}02 \cdot 30 = 49{,}4°C$. De la expresión de la temperatura se concluye que el tiempo para el que se alcanza es: $t_S = 3{,}91A$; para el caso de una especificación más rígida en cuanto a tolerancias (<0,1%), tenemos que el tiempo de establecimiento del sistema sería el que corresponde a que la salida alcance el valor de $50 - 0{,}001 \cdot 30 = 49{,}97°C$ lo que ocurre para un tiempo de establecimiento mucho mayor, $t_S = 6{,}91A$.

Como se puede observar, el tiempo de establecimiento aumenta en la medida que crece la especificación de exactitud del sistema.

Resumen

- La instrumentación comprende todas las técnicas, equipos y metodologías relacionados con el diseño, construcción y aplicación de dispositivos físicos para mejorar, completar y aumentar la eficiencia de los mecanismos de percepción del ser humano.
- La instrumentación electrónica es la técnica que se ocupa de la medición de cualquier tipo de magnitud física, de la conversión de la misma a magnitudes eléctricas y de su tratamiento para proporcionar información al sistema de control, al operador o a ambos.
- Un sistema que mantiene una relación entre la salida y la entrada de referencia, comparándolas y usando la diferencia como medio de control, se denomina sistema de control realimentado.
- En un sistema de control en lazo abierto la salida del proceso no afecta la acción de control. En un sistema en lazo cerrado la salida se realimenta al sistema de control.
- La instrumentación es útil en el control, porque proporciona las técnicas y los procedimientos necesarios para medir la variable de proceso e informar al sistema de control.
- El ámbito de la instrumentación en un sistema de control de procesos incluye, fundamentalmente, el sistema de medida integrado por el sensor y el acondicionamiento de señal.
- La curva de calibración se define, en general, por una ecuación o por una tabla de valores. En particular, las curvas de calibración lineales se definen mediante un punto, la sensibilidad y el error de no linealidad.
- Los límites de la curva de calibración se definen mediante el campo de medida, alcance y la salida a fondo de escala.
- Otros valores que pueden resultar interesantes para definir la curva de calibración son la zona muerta, la resolución, la histéresis, las derivas y la zona de saturación.
- La veracidad mide el grado de concordancia entre los resultados de una medida y el valor exacto mientras que la precisión mide el grado de concordancia de los resultados entre sí. El término exactitud se emplea para referirse a ambos términos conjuntamente.
- La precisión se determina a partir de la repetibilidad y la reproducibilidad.
- Las formas más habituales para expresar la exactitud de un sensor o de un instrumento de medida son el error en tanto por ciento sobre el fondo de escala y en tanto por ciento sobre la lectura.

- La calibración permite disminuir los errores sistemáticos. Los procedimientos de calibración más elementales son la calibración a un punto y la calibración a un punto y la sensibilidad.
- El comportamiento de los sistemas de instrumentación presenta un retardo en su respuesta de salida a partir de los estímulos de entrada.
- La caracterización de la forma de la respuesta temporal o frecuencial de un sistema de instrumentación permite definir su respuesta dinámica.
- El ancho de banda y la distorsión armónica total son las principales características dinámicas en el dominio de la frecuencia.
- El tiempo de establecimiento y la sobreoscilación son la principal forma de caracterizar la respuesta temporal de un sistema de instrumentación.

Problemas propuestos

1.1 Amplíe el sistema de control del tanque de agua de la figura 1.7 y considere una válvula de seguridad que deberá actuar si por cualquier circunstancia no prevista el tanque se desborda.

1.2 Construya el diagrama de bloques de un sistema para controlar la velocidad de un vehículo. ¿Es un sistema realimentado?

1.3 Construya el diagrama de bloques de un sistema para controlar la dirección de un vehículo. ¿Es un sistema realimentado?

1.4 ¿Por qué es importante la instrumentación en el control de procesos?

1.5 Se dispone de un sensor de presión que presenta la siguiente curva de calibración:

	Entrada, x_i. Presión (kPa)										
	0,0	10,0	20,0	30,0	40,0	50,0	60,0	70,0	80,0	90,0	100,0
Salida, y_i (V)	0,005	0,098	0,224	0,300	0,405	0,520	0,602	0,715	0,799	0,902	0,999

a) Linealizar la curva de calibración por el método de mínimos cuadrados.
b) Determinar el error de linealidad expresado en % sobre el fondo de escala.

Nota: la recta resultante aplicando la aproximación por mínimos cuadrados, $y = a + bx$, se puede determinar teniendo en cuenta que:

$$a = \frac{\sum_{i=1}^{n} y_i - b \cdot \sum_{i=1}^{n} x_i}{n} \qquad b = \frac{n\sum_{i=1}^{n} x_i \cdot y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n \cdot \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

donde n es el número de medidas realizadas.

1.6 Se dispone de dos sensores de temperatura para medir entre 0 y 100°C. El primero proporciona una tensión de salida $v_1(T) = 50 \cdot T\ \mu V$ y el segundo, $v_2(T) = 100 \cdot T\ \mu V$, donde T es la temperatura ex-

presada en °C. Suponiendo que el instrumento para medir la tensión presenta un error de ±10 μV. ¿Con cuál de los dos sensores se podría determinar la temperatura con mayor exactitud?

1.7 Las hojas de características de un acelerómetro indican:

Campo de medida:	±50g
No linealidad:	0,2% respecto del fondo de escala
Tensión de salida a 0g:	2,50 ± 0,35 V
Sensibilidad:	40,0 ± 5,0 mV/g

a) Deteminar la curva de calibración.
b) Determinar el máximo error en % sobre el fondo de escala si no se calibrase el sistema.
c) Deteminar el máximo error en % sobre el fondo de escala si se calibrase perfectamente el cero y la ganancia.
d) Suponiendo que se va a realizar la calibración a partir de la salida a 0g y a 1g (valores muy fácilmente obtenibles, puesto que 1g es la aceleración de la gravedad), proponer un procedimiento de calibración.

1.8 Sea un sistema definido por su función de transferencia $G(\omega) = (\omega^2 + 15\omega + 35)/(\omega^3 + 10\omega^2 + 10)$. Representar de forma aproximada el módulo de la ganancia en decibelios en función del logaritmo de la frecuencia ω.

1.9 En el problema anterior, representar la fase de $G(\omega)$ en función de la frecuencia ω.

1.10 Dibuje el diagrama de Bode de una red LC con L = 10 μH y C = 50 nF, considerando que se introduce una tensión de entrada v_i y que se obtiene la tensión de salida v_o sobre el condensador.

1.11 En el problema anterior, ¿qué ocurre con la tensión de salida cuando se igualan los valores de la impedancia capacitiva y la inductiva?

1.12 Si en el problema 1.10 se añade una resistencia de 1K en serie con la bobina, ¿cómo se modifica la respuesta del conjunto?

1.13 Sea un sistema con una función de transferencia $G(j\omega) = 100/(1000+j\omega)$. ¿Qué ganancia presenta en continua? ¿Cuál es el diagrama de Bode de esa función de transferencia?

1.14 Si introducimos una señal senoidal de 500 Hz en el sistema del problema anterior, ¿cuál es la salida que se obtiene?

1.15 ¿Qué forma de onda de tensión de salida aparece en el sistema del problema 1.13 si la entrada es una señal cuadrada de 3 V de pico a pico y 400 Hz de frecuencia.

2

Amplificación

2.1 Conceptos generales
2.2 El amplificador operacional

2.1 Conceptos generales

Una etapa de amplificación en un sistema electrónico es un bloque con una entrada y una salida que se encarga de tomar la señal presente en esa entrada y situarla en la salida con un mayor nivel de potencia.

En el presente capítulo estudiaremos las particularidades de los amplificadores y justificaremos su necesidad en las distintas etapas de un sistema de instrumentación.

Las señales eléctricas pueden ser corrientes o tensiones por lo que tendremos cuatro tipos de amplificadores según sean su entrada y su salida:

a) Amplificadores con entrada y salida en tensión.
b) Amplificadores con entrada y salida en corriente.
c) Amplificadores con entrada en tensión y salida en corriente.
d) Amplificadores con entrada en corriente y salida en tensión.

La ganancia de potencia A_p se define como la potencia de la señal de salida P_o dividida entre la potencia de la señal de entrada P_i:

$$A_p = \frac{P_o}{P_i} \tag{2.1}$$

Si se pretendiese obtener una ganancia elevada, se debería tener un sistema que absorbiese baja potencia, es decir, cuando la señal de entrada se comporte como una fuente de corriente, sería un cortocircuito mientras que para entradas en tensión, el equivalente de entrada debería ser un circuito abierto (figura 2.1a).

En la parte de salida, si pretendemos que la variable de salida no se vea afectada por la carga sobre la que se desarrolla la potencia, cuando la salida sea en tensión la impedancia de salida interesaría que fuese nula y que, cuando la salida sea en corriente, infinita (figura 2.1b).

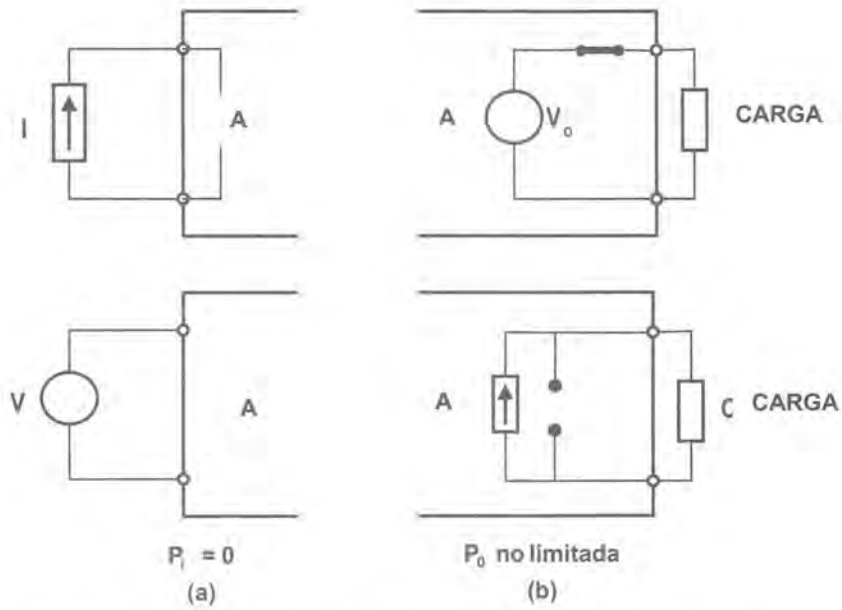


Figura 2.1. Equivalentes de entrada (a) y de salida (b) de amplificadores ideales.

En estas condiciones se conseguiría una óptima amplificación de potencia sin perder señal en ningún punto puesto que el amplificador no precisa potencia en la entrada y toda cuanta genera en la señal de salida es entregada sobre la carga.

Las situaciones de impedancia nula o infinita planteadas no son reales y, en la práctica, siempre se tiene algún nivel de impedancia lo que por otra parte resulta imprescindible a la entrada del amplificador para representar la influencia de la lectura de la señal de entrada, es decir, si no se toma ninguna energía de la entrada, no será posible amplificarla. Si, a ello añadimos la inevitable presencia de impedancias de salida en la fuente de señal, resultan esquemas como los de la figura 2.2 donde se representan los equivalentes de Thévenin (a) y Norton (b) de la fuente de señal que introducimos en el amplificador.

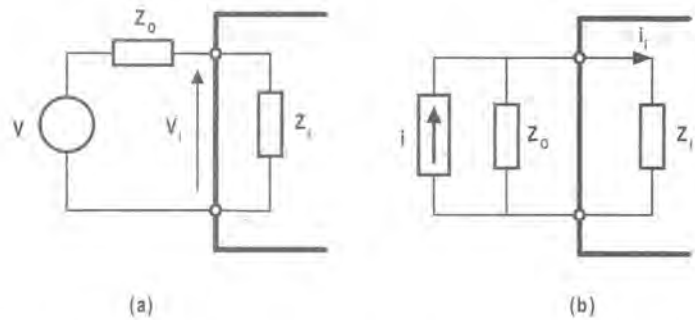


Figura 2.2. Conexión de una fuente de señal a la entrada de un amplificador: (a) equivalente Thévenin; (b) equivalente Norton.

En estos casos se producen efectos de carga que provocan la aparición de tensiones o corrientes de entrada menores que las que proporcionarían las fuentes en vacío (cuando es fuente de tensión) o en cortocir-

cuito (cuando es fuente de corriente) y que vienen determinadas por el correspondiente divisor resistivo. Para el caso de la figura 2.2a, la tensión en los bornes de entrada del amplificador es:

$$v_i = v\frac{Z_i}{Z_i + Z_0} < v \tag{2.2}$$

Mientras que para el caso de la figura 2.2b, tenemos que la corriente de entrada al amplificador sería:

$$i_i = i\frac{Z_o}{Z_i + Z_0} < i \tag{2.3}$$

quedando de manifiesto el efecto de la carga como un error sistemático en la lectura que depende del valor de Z_o y de Z_i y que podría ser corregido. Este efecto se manifestaría también en la propia salida del amplificador que poseerá una impedancia salida finita y distinta de cero ocasionando una nueva pérdida de señal sobre la carga.

Por otra parte, se puede demostrar que la impedancia de entrada del amplificador no influye en la relación S/N aunque sí que tiene influencia la impedancia de salida de la fuente de señal. Los circuitos genéricos de la figura 2.3 muestran una fuente de tensión de señal en el caso (a) y una fuente de corriente de señal en el caso (b) y, para ambos, dos fuentes de ruido i_n y v_n; se puede demostrar que la relación S/N no se ve afectada por la impedancia de entrada del amplificador Z_i y aumenta con la idealidad de las fuentes de señal (Z_o baja para la fuente de tensión y alta para la de corriente) según:

$$S/N = \frac{v_s^2}{(v_n + Z_o i_n)^2} \text{ (para 2.3a)} \qquad S/N = \frac{Z_o^2 i_s^2}{(v_n + Z_o i_n)^2} \text{ (para 2.3b)} \tag{2.4}$$

para los circuitos correspondientes de la figura 2.3. En condiciones ideales ($Z_o = 0$ en el primer caso y $Z_o = \infty$ en el segundo) la relación S/N (2.4) se transforma, respectivamente, en $S/N = (v_s/v_n)^2$ y $S/N = (i_s/i_n)^2$

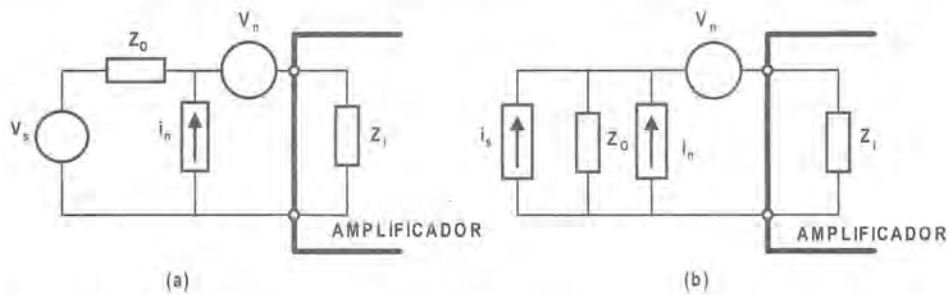


Figura 2.3. Inclusión de las fuentes de ruido en las entradas de los amplificadores con equivalente Thévenin (a) y equivalente Norton (b).

Las conclusiones de las expresiones (2.4) se pueden aplicar a la etapa de salida del amplificador, en el cual la fuente de señal es el propio amplificador y se está entregando la potencia sobre otro sistema: desde el punto de vista de la relación señal/ruido (S/N), la salida del amplificador debe parecerse lo más posible a una fuente ideal, no influyendo la carga sobre la relación S/N.

En cualquiera de los anteriores casos, si la señal de partida ya incluye el ruido, el amplificador se limitará a amplificar el conjunto no produciendo —en condiciones ideales— ningún efecto adicional.

Así pues, con referencia a las distintas arquitecturas de los sistemas de instrumentación que se han estudiado en los primeros apartados y como quiera que el objetivo suele ser llevar información desde un proceso

(desde la salida del sensor) a otro punto, nos podríamos preguntar por la necesidad de amplificar: en el fondo, si sólo estamos incrementando los niveles, podríamos dejarlo como está y presentar la información entre dos valores más bajos.

Por ejemplo, si un sensor proporciona una señal comprendida entre 0 y 1 mV, lo único que necesitamos es un instrumento que realice la representación oportuna con ese fondo de escala. ¿Para qué necesitaríamos levantar el nivel de la señal hasta llevarla a 0 – 10 V? La razón está en que si la introducción de ruido tiene probabilidad de ocurrir en cualquier punto del circuito, su importancia relativa será menor cuanto mayor sea el nivel de la señal; así, si el ruido equivaliese a una tensión adicional de 0,1 mV, con la señal de 1 mV, la relación S/N resultaría ser de 20 dB mientras que si la señal fuese de 1 V, S/N = 80 dB. Supongamos que el ruido se introduce con la misma potencia en todas las zonas del circuito (figura 2.4): se puede ver cómo el efecto del ruido va teniendo menos importancia a medida que se avanza en el circuito y la señal se va amplificando.

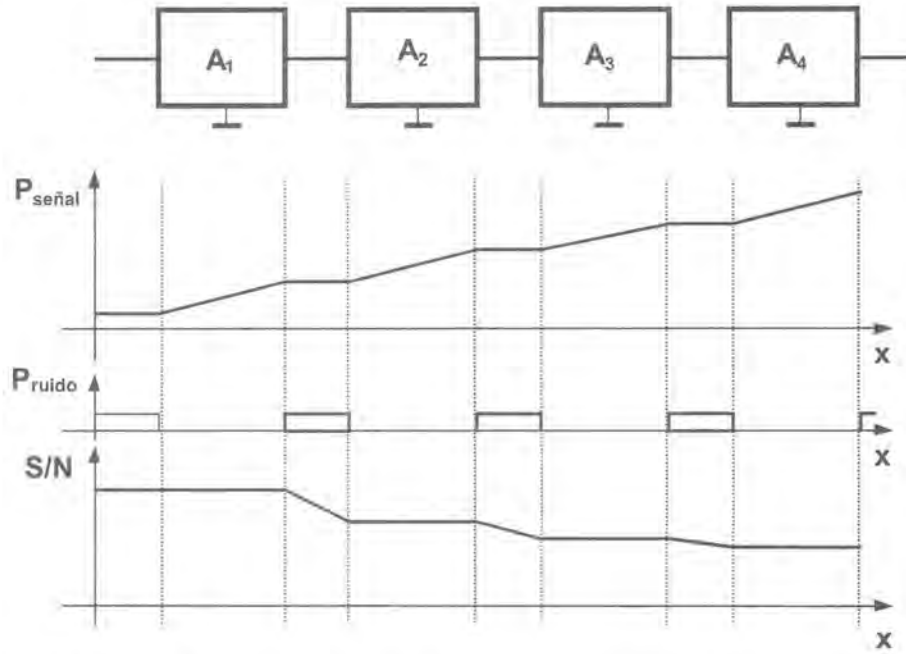


Figura 2.4. Pérdida de valor de la relación S/N a medida que se avanza en la cadena de amplificación o tratamiento. No obstante, la pérdida cada vez es menor ya que la potencia de señal cada vez es más grande.

A modo de conclusión, cuanto antes podamos elevar los niveles de las señales, tanto mejor porque los posibles ruidos que se introducirán en el circuito resultarán cuantitativamente menores: si no tienes muy claro qué hacer con una señal, ¡amplifícala!

De esta forma, observamos que el uso de amplificadores resulta trascendental en cualquier sistema pero, en particular en el caso de la instrumentación electrónica puesto que la relación S/N cobra una gran importancia.

Pero los amplificadores no presentan características ideales, sino que tienen desviaciones en su funcionamiento, introduciendo una serie de limitaciones de carácter estático y dinámico definidas por los valores de los parámetros ya estudiados en el capítulo anterior (apartados 1.1.4 y 1.1.5). Así, un amplificador presenta una serie de no idealidades:

- Ganancia no lineal.
- Generación interna de ruido.
- Errores estáticos producidos por desviaciones.
- Distorsión armónica y distorsión de fase provocadas por un ancho de banda limitado.

Según el uso que se vaya a hacer del amplificador se tendrán o no en cuenta estas no idealidades y, en el más sencillo de los casos se podrá considerar que es ideal, quedando caracterizado por una serie de valores que definen su comportamiento y que son, en una primera aproximación:

a) La ganancia de potencia, A_p, parámetro general que hace referencia al efecto básico del amplificador.
b) Las ganancias de tensión, A_v, de corriente, A_I, de trans-resistencia R o de trans-conductancia G, según sea la naturaleza de las señales de entrada y salida.
c) Las resistencias de entrada y salida.

En una segunda aproximación, hay que considerar posibles desviaciones en este funcionamiento ideal tales como:

a) La presencia de efectos de desviación (*offset*) en el valor de salida, introduciendo el correspondiente error en el valor de salida.
b) La presencia de ganancia no-lineal.
c) Los cambios en la ganancia en función de la frecuencia, debidos a la propia ganancia, a la presencia de impedancias de entrada y salida o a efectos de carga.

En una tercera aproximación y para un estudio más completo, se pueden considerar otros efectos no deseados como el ruido generado internamente y que se añade al que poseía la propia señal a la entrada empeorando la relación S/N.

En resumen, en la figura 2.5, se pueden observar estos efectos sobre una señal que pretendía ser amplificada.

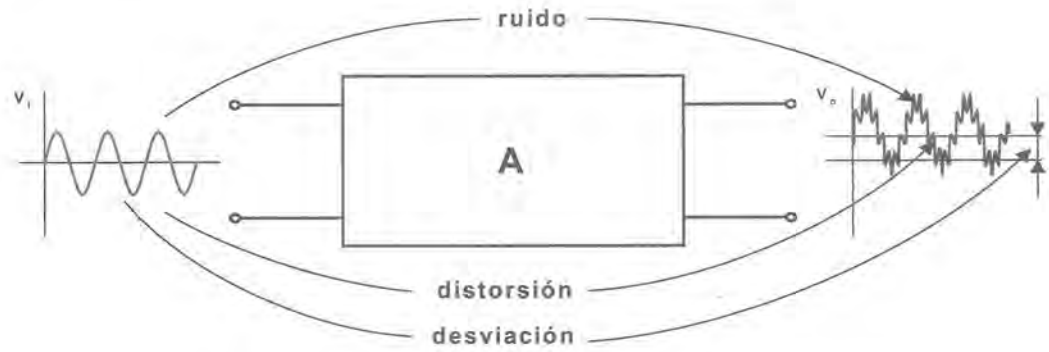


Figura 2.5. Efectos no deseados sobre las señales que son amplificadas.

El panorama que se ha presentado parece ser bastante desalentador y puede hacer pensar que cualquier sistema funciona mejor que uno electrónico. Sin embargo, la realidad es que la importancia de esos efectos debe ser cuantificada y —en muchas ocasiones— resulta despreciable por lo que los amplificadores suelen cumplir su función de una forma digna.

En el caso particular de la Instrumentación Electrónica, con señales débiles, relaciones S/N bajas, frecuencias altas en muchos casos, el diseño se pone difícil...

2.1.1 AMPLIFICAR Y MODIFICAR LOS VALORES DE UNA SEÑAL

La amplificación puede resultar muy útil para proporcionar una ganancia rápida de potencia a una señal y conseguir que se vea menos afectada por el ruido y demás fenómenos adversos. Pero las etapas de amplificación se pueden usar para otros fines diferentes aunque siempre ligados a los cambios en los valores de las magnitudes —tensión y corriente— que definen una determinada señal.

Estos casos se dan en muchos diseños en que se precisa una señal con un margen de variación diferente al que posee. Los circuitos que realizan esa función son circuitos de acondicionamiento de señal cuya misión es asignar de forma unívoca un nuevo conjunto de valores a los actuales (figura 2.6):

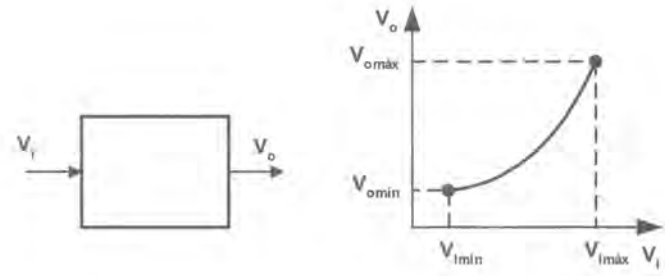


Figura 2.6. Circuito general de acondicionamiento de señal.

La función del circuito de la figura 2.6 resulta muy general ya que en la mayoría de los casos, la señal es modificada de unas formas muy concretas:

a) Se produce una adaptación lineal de niveles siguiendo una expresión tal como $v_0 = av_i + b$ constituyendo un circuito adaptador de nivel.

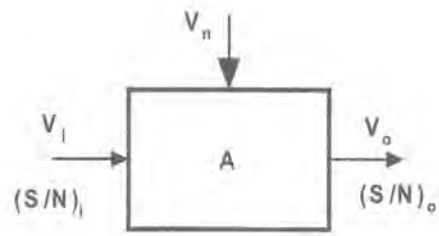
b) Se produce una atenuación lineal de la señal $v_0 = av_i$ con $a < 1$ que define el comportamiento de un atenuador.

c) Se produce una detección no lineal de nivel provocando una salida digital que constituye el funcionamiento de un comparador básico.

Las dos primeras aplicaciones son de tipo analógico mientras la segunda implica el paso a señal digital puesto que todas las situaciones se codifican en función de dos estados únicamente.

2.1.2 RUIDO GENERADO

La relación S/N es uno de los parámetros más importantes siempre que hablemos de sistemas de instrumentación ya que acota su incertidumbre. Como hemos visto, las señales incorporan niveles de ruido y, además, se va añadiendo más ruido a lo largo de la cadena de tratamiento de los datos, de tal forma que poco a poco va disminuyendo la relación S/N.



$(S/N)_o < (S/N)_i$

Figura 2.7. Modelo de inclusión de ruido en un amplificador.

Los amplificadores —como cualquier circuito electrónico— no son una excepción a la hora de incrementar los niveles de ruido y cualquier señal que los atraviese incrementará su nivel de ruido disminuyendo la relación S/N (figura 2.7). Este ruido añadido por el amplificador, v_n provocado por los distintos componentes electrónicos que lo constituyen (resistencias, diodos y transistores) [2.18] se conoce como ruido interno.

El conjunto del ruido interno generado por el amplificador se puede cuantificar mediante un parámetro global conocido como figura de ruido (*Noise Figure* NF) definido como el cociente entre la relación S/N a la entrada y la relación S/N a la salida:

$$NF = 10\log\frac{(S/N)_i}{(S/N)_o} \tag{2.5}$$

expresión que siempre resulta mayor que cero y que presenta valores elevados en amplificadores ruidosos y bajos en amplificadores poco ruidosos. No obstante, la importancia del ruido generado es mayor cuando las señales son más débiles lo que significa que las primeras etapas de amplificación resultan críticas en este aspecto (figura 2.4), mientras que en las etapas finales en donde los niveles de potencia son más elevados, el ruido generado internamente puede despreciarse.

Cuando diseñemos las etapas de entrada que deben aportar la amplificación para señales débiles procedentes de un sensor, deberemos prestar especial atención a la selección de todos sus componentes para tratar de minimizar el ruido interno ya que su magnitud podría llegar a ser comparable con el de la señal.

Este análisis del ruido generado resulta una primera aproximación al problema puesto que hay más factores a tener en cuenta, que permiten un mejor tratamiento del asunto y que se tratarán en el apartado 2.2.3.3.

2.1.3 FILTRADO

Ya hemos visto que cualquier circuito electrónico posee un determinado comportamiento en frecuencia que puede introducir modificaciones en el espectro de la señal de entrada produciendo distorsión armónica; en muchas ocasiones, este hecho resulta no deseable y se buscarán circuitos de banda ancha que permitan el paso de todas las componentes de frecuencia de la señal de entrada sin introducir atenuaciones en ninguna de ellas.

Sin embargo, en otras ocasiones se busca limitar la banda pasante para eliminar o reducir todas las componentes excepto unas en concreto. Cuando el circuito está diseñado con este criterio estamos en presencia de un filtro y se conoce como filtrado la acción de selección de determinadas frecuencias y la de eliminación o atenuación de otras.

Para comprender la necesidad de filtrado de una determinada señal hay que tener en cuenta lo siguiente:

a) El ruido presente en cualquier sistema suele tener un espectro de frecuencias muy amplio con lo que su contribución a la señal será mayor en la medida que el ancho de banda del sistema sea mayor.

b) Cualquier señal posee un espectro de frecuencias que incorpora datos necesarios y algunos que no lo son. Desde el punto de vista de instrumentación todo lo que forma parte de la información que se pretende tratar constituye la señal en sentido estricto mientras que el resto, sea cual sea su origen, puede ser considerado como "ruido".

Como lo que perseguimos es obtener la información de un determinado proceso, debemos tener las señales sin presencia de componentes no deseadas y limitar el ancho de banda a las componentes que nos interesan, atenuando todas las demás que no nos proporcionan ninguna información relevante y sólo contribuirían a reducir la relación S/N según se muestra en la figura 2.8.

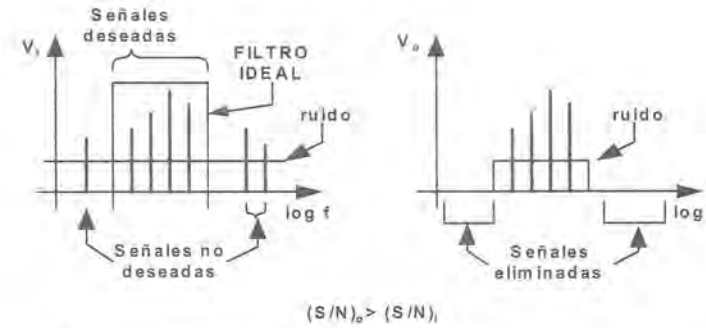


Figura 2.8. Mejora de la relación S/N por reducción del ancho de banda.

Así pues, reducir adecuadamente el ancho de banda de los sistemas mejora la relación S/N.

2.2 El amplificador operacional

Existen muchos circuitos electrónicos capaces de amplificar una señal de tensión o de corriente y desarrollar más potencia sobre una carga que la que podría producir la señal original. En muchos textos de electrónica general se dedican amplios capítulos a todos estos circuitos basados en el uso de transistores discretos, sean de tipo bipolar o de tipo FET, pero en el campo de la instrumentación electrónica el dispositivo amplificador por excelencia es el amplificador operacional o, como se le suele conocer, el operacional.

No cabe duda que el amplificador operacional está constituido por etapas, formadas, a su vez, por diversos transistores. Si bien su estudio interno tiene interés intrínseco, aquí lo asumiremos como un componente electrónico más, estudiaremos sus características externas en detalle y veremos cómo afectan a su funcionamiento, a su ámbito de aplicación y al diseño de los circuitos en los que esté inmerso, dejando para textos más específicos el análisis pormenorizado de su estructura interna [2.8].

2.2.1 EL AMPLIFICADOR OPERACIONAL IDEAL

Las excelentes prestaciones del amplificador operacional —operacional en lo sucesivo— le han proporcionado un inmenso campo de aplicación dentro del mundo de la electrónica analógica constituyendo el bloque básico de la mayoría de los amplificadores.

Estas características se acercan mucho a las que se podrían pedir de cualquier amplificador ideal, de tal manera que en gran número de aplicaciones puede ser considerado como ideal sin introducir demasiados errores en el cálculo y, en cualquier caso, siempre será la primera aproximación dentro de un proceso de diseño.

El operacional se representa con un símbolo como el que se muestra en la figura 2.9 con dos entradas, una inversora (–) y otra no inversora (+), una salida y dos terminales de alimentación.

La función que desarrolla el operacional es la de amplificar con una ganancia de tensión A_d la tensión diferencial v_d según:

$$v_o = A_d v_d \tag{2.6}$$

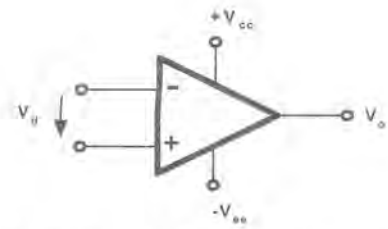


Figura 2.9. Símbolo del amplificador operacional.

La consideración de idealidad del operacional queda definida por los valores extremos de los diversos parámetros que lo definen y que son los de cualquier amplificador:

- Resistencia de entrada infinita, lo que obliga a que la corriente de entrada por cualquiera de sus dos entradas sea siempre nula.
- Resistencia de salida nula, lo que implica infinita capacidad de proporcionar corriente sobre cualquier carga.
- Ganancia de tensión A_d infinita, lo que significa que si su salida es finita, la tensión diferencial de entrada deberá ser nula.

Como consideración adicional, hay que tener en cuenta que la tensión de salida siempre se encuentra limitada por las tensiones de alimentación y no podrá sobrepasarlas en ningún caso con lo que, si el operacional está alimentado entre un valor positivo $+V_{cc}$ y un valor negativo $-V_{ee}$, se cumplirá que la tensión v_o siempre estará comprendida entre ellas, es decir:

$$-V_{ee} \leq v_o \leq +V_{cc} \tag{2.7}$$

Las consideraciones de idealidad permiten un estudio simplificado del operacional y de los circuitos construidos en torno a él, obteniendo algunas aplicaciones muy interesantes en los circuitos de tratamiento de señal. Cabe hablar de aplicaciones analógicas en los casos en que la salida del amplificador produce tensiones de salida que pueden tomar cualquier valor intermedio entre dos dados con las limitaciones impuestas por la expresión (2.7) y otras aplicaciones digitales en las que sólo tomará dos estados sin posibilidad de ningún valor intermedio.

En la tabla 2.1 se presenta un resumen de las principales aplicaciones analógicas del operacional ideal constituyendo algunos de los circuitos más habituales para tratamiento de señal. En la tabla 2.2 se resumen aplicaciones digitales en que la salida del amplificador puede tomar sólo dos valores.

Las aplicaciones incluidas en la tabla 2.1 se han obtenido teniendo en cuenta la idealidad del operacional que implica que cuando está realimentado negativamente y no saturado, la tensión diferencial de entrada es nula. Llegar a obtener cualquiera de esas expresiones resulta muy simple y aquí omitiremos la demostración matemática aunque el lector interesado puede acudir a cualquier texto de electrónica analógica de nivel básico [2.6], [2.10], [2.17].

Existen muchas más aplicaciones de los amplificadores operacionales tanto de tipo analógico como digital y una multitud de circuitos en los que se combinan con otros bloques o circuitos para realizar funciones más complejas. El número de casos que podemos encontrar resulta tan amplio que resultaría tedioso repetirlos aquí ya que forman parte de cualquier texto de electrónica analógica o de electrónica general. En ellos se pueden encontrar circuitos con amplificación no lineal (logarítmica o antilogarítmica), multiplicadores y divisores, recortadores, limitadores, osciladores, detectores, etc. [2.1] constituyendo un amplio abanico de posibilidades para el diseñador.

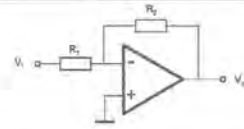
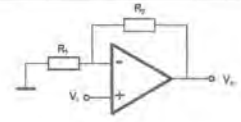
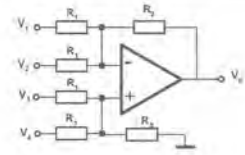
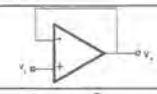
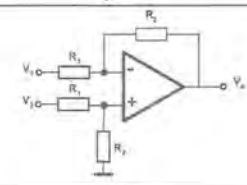
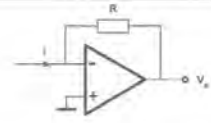
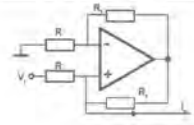
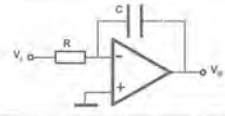
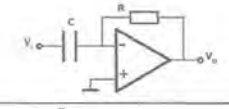
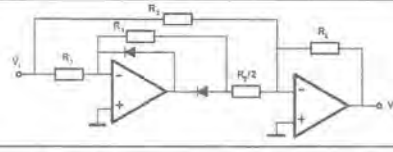
Nombre	Circuito	Aplicaciones y comentarios
Amplificador inversor $v_0 = -\frac{R_2}{R_1} v_i$		Ganancia negativa con cualquier valor absoluto.
Amplificador no inversor $v_0 = \frac{R_1 + R_2}{R_1} v_i$		Ganancia positiva siempre mayor que 1.
Amplificador genérico $v_0 = \frac{R_2}{R_1}(v_3 + v_4 - v_2 - v_1)$		Es un sumador/restador con ganancia. Se pueden hacer sumadores y restadores eliminando algunas entradas.
Seguidor de emisor o seguidor de tensión $v_0 = v_i$		No implica carga para la fuente de señal y admite cualquier carga en su salida.
Amplificador diferencial $v_0 = \frac{R_2}{R_1}(v_2 - v_1)$		Es un caso particular del amplificador genérico de utilidad en instrumentación.
Convertidor corriente-tensión $v_0 = -R\,i$		Circuito básico para obtener una salida de tensión correspondiente a una entrada en corriente.
Convertidor tensión-corriente $i_0 = v_i / R$		Circuito básico para obtener una salida de corriente correspondiente a una entrada en tensión.
Integrador $v_0 = -\frac{1}{RC}\int v_i(t)\,dt$		Circuito para calcular la integral de una señal de tensión de entrada.
Derivador $v_0 = -RC\,\frac{dv_i(t)}{dt}$		Circuito para calcular la derivada de una señal de tensión de entrada.
Rectificador de precisión $v_0 = \lvert v_i \rvert$		Circuito de precisión, capaz de rectificar eliminando la tensión de codo del diodo.

Tabla 2.1. Circuitos analógicos con amplificadores operacionales y función que realizan considerando que su comportamiento es ideal.

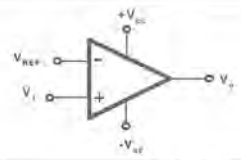
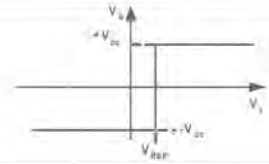
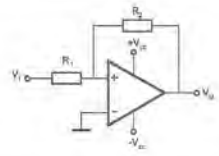
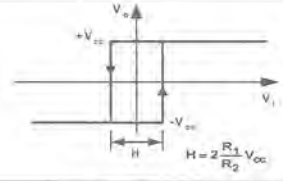
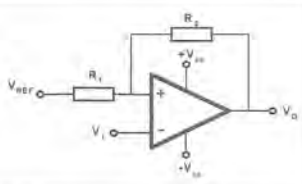
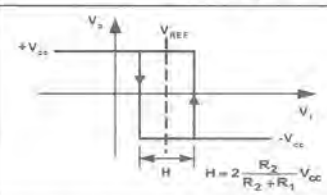
Nombre	Circuito	Función	Aplicaciones y comentarios
Comparador básico			Circuito básico de comparación con una tensión referencia.
Comparador con histéresis			Circuito para evitar los cambios múltiples en los instantes de cambio del comparador mediante una tensión de histéresis H.
Comparador con histéresis II			Como el anterior pero la histéresis tiene lugar alrededor de la tensión de referencia.

Tabla 2.2. Circuitos de comparación con amplificadores operacionales y función que realizan considerando que su comportamiento es ideal.

Sin embargo, en el ámbito de la instrumentación electrónica sólo se usa un reducido grupo de circuitos que usan el operacional como núcleo y que son aquellos capaces de proporcionar amplificación o de cambiar niveles y/o magnitudes (tabla 2.1) o detectar niveles de una señal (tabla 2.2).

A pesar de la aparente sencillez de las aplicaciones, al considerar que el operacional no es ideal sino real, aparecen una serie de efectos secundarios que pueden llevar a graves errores si se usan los circuitos tal y como están presentados en esas tablas. Así, podemos decir que las tensiones representadas en las fórmulas de la tabla 2.1 no son exactas o, en algunos casos, no se corresponden en absoluto con la realidad si se considera que el operacional empleado es real.

Los efectos secundarios en los amplificadores operacionales debidos a su no-idealidad conducen a la aparición de términos adicionales no deseados que constituyen un error permanente en la salida del circuito o en la introducción de distorsión armónica o de fase no deseada de la señal.

Todos estos efectos pueden tener más o menos importancia en función de su cuantía por lo que no siempre deben ser tenidos en cuenta; en general, resultan valores pequeños en términos de valor absoluto y en la mayoría de las aplicaciones analógicas podrían ser despreciados sin detrimento de la operatividad del diseño. En realidad, sólo son importantes cuando su efecto es comparable con los niveles de señal que se manejen.

Desafortunadamente, es el caso de las señales en instrumentación...

2.2.2 EL AMPLIFICADOR OPERACIONAL REALIMENTADO EN TENSIÓN

En este apartado daremos solución al problema planteado en el apartado anterior para el operacional más habitual: el que se realimenta en tensión y que actúa como un amplificador de tensión, es decir, se caracteriza por:

- Una elevada resistencia de entrada.
- Una baja resistencia de salida.
- Una ganancia de tensión muy elevada.

Si recordamos lo que decíamos al comienzo del apartado 2.1.2, las especificaciones se parecen a las de un amplificador ideal excepto en que en aquel caso, los valores eran extremos: ahora no son infinitas ni la resistencia de entrada ni la ganancia y la resistencia de salida no es nula. En las siguientes secciones veremos cómo afectan estos valores y la importancia que tienen otros factores adicionales que pueden dar al traste con cualquier diseño.

La diferencia de comportamiento del operacional respecto al ideal debe ser cuantificada. El resultado de esa cuantificación es un error que puede ser asumible en muchos casos, resultando despreciable la no-idealidad del dispositivo, mientras que, en otros, no será posible considerar al operacional como ideal y habrá que tratar de reducir los efectos de esas no-idealidades o recurrir a utilizar otro dispositivo que aporte menos errores. Para ello, será imprescindible conocer en detalle cada uno de los parámetros reales de los operacionales.

2.2.3 PARÁMETROS REALES DE LOS OPERACIONALES

Vamos a ir analizando cada uno de los parámetros que definen el funcionamiento de un operacional real, aprendiendo a cuantificar su efecto sobre la salida, otorgándole la importancia relativa que le corresponda y analizando cuáles son más críticos por su influencia en cualquier aplicación.

Para entender las causas de la no-idealidad consideremos un operacional cualquiera, por ejemplo el clásico 741 que es uno de los más utilizados (ver CD adjunto). Se trata de un dispositivo de tecnología bipolar y de uso general (*general purpose*) que no tiene unas prestaciones excepcionales pero que permite ilustrar todos los parámetros reales que definen el funcionamiento de cualquier otro.

En la figura 2.10 se muestra su símbolo de circuito, sus pines de conexión y su estructura interna.

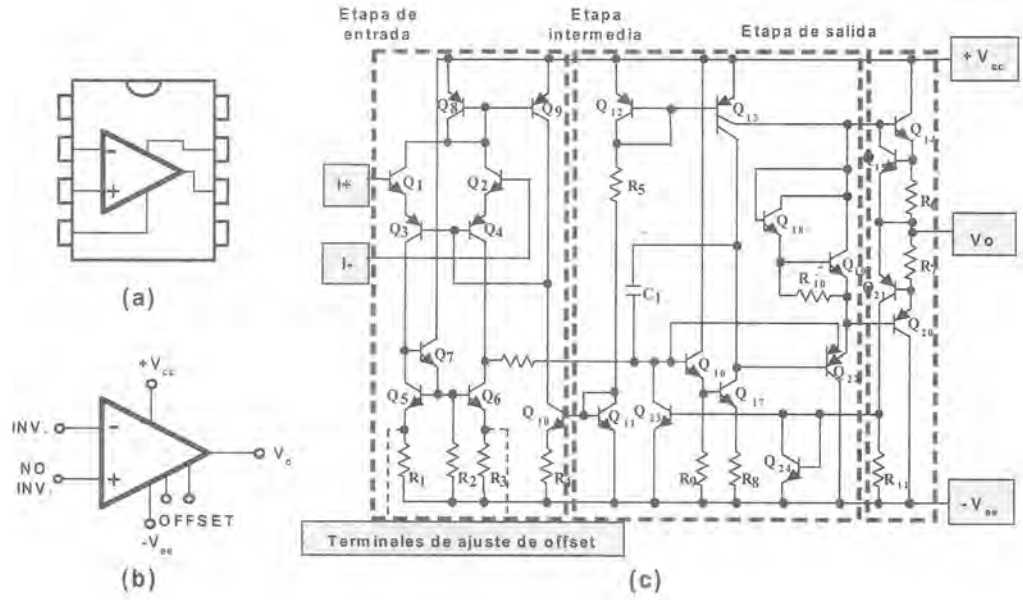


Figura 2.10. Amplificador operacional 741: (a) patillaje; (b) símbolo de circuito y, (c) estructura interna.

Aunque no entraremos a describir detalladamente cómo funciona [2.8], sí que describiremos someramente su estructura, compuesta por tres etapas principales:

- Etapa de entrada de tipo diferencial y que aporta una buena parte de la ganancia total del circuito [2.8].
- Etapa intermedia que se encarga fundamentalmente de la adaptación de niveles [2.8] ya que el acoplamiento entre etapas se hace en continua.
- Etapa de salida que proporciona la potencia necesaria para manejar pequeñas cargas y que incorpora protecciones [2.8].

Haciendo un recorrido por su estructura desde las entradas a la salida nos vamos encontrando con las causas que provocan un comportamiento que difiere del funcionamiento ideal antes descrito. Comencemos pues, por la etapa de entrada:

La etapa de entrada del 741

Observamos una clásica entrada diferencial bipolar con las dos entradas conectadas a masa; para su funcionamiento precisa corrientes en las bases de los transistores Q_1 y Q_2, de tal manera que **no es cierta la suposición ideal de corrientes nulas en las entradas** (figura 2.11).

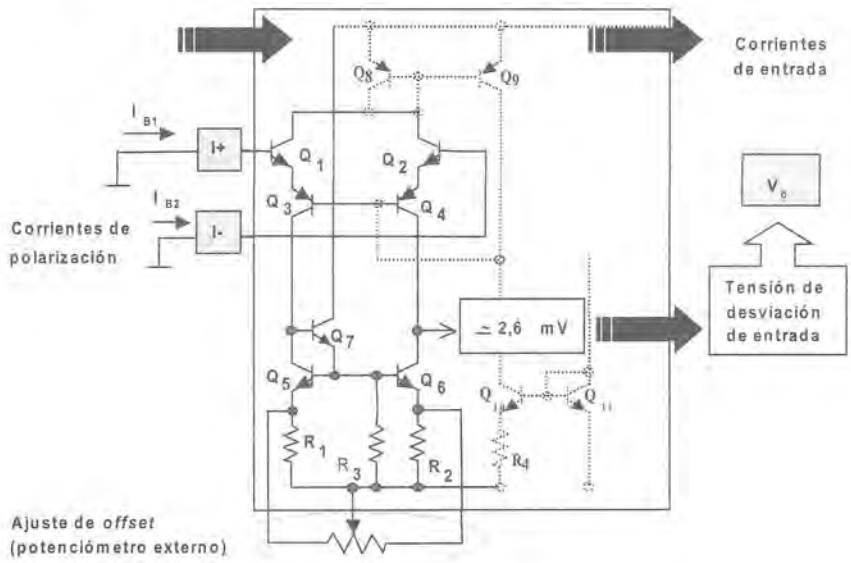


Figura 2.11. Etapa de entrada del 741.

Así, aparecen **corrientes de polarización** I_{B1} e I_{B2} que serían iguales si las dos ramas de la etapa diferencial lo fueran. Como quiera que eso resulta imposible en la práctica (aunque sí resultan muy parecidas no son estrictamente iguales) las dos corrientes de polarización (*bias current*) serán distintas. No es posible saber cuál de las dos va a ser mayor así que se define un parámetro que recoge la diferencia entre ambas, denominado **corriente de desviación** (*offset current*) y que se define como:

$$I_{io} = I_{B1} - I_{B2} \tag{2.8}$$

De la misma forma, la presencia de asimetrías en las dos ramas de la etapa diferencial provocará —incluso con corrientes de polarización hipotéticamente iguales— una tensión en la salida (recordemos que, con las dos entradas a masa, la tensión de salida debería ser nula). Esa tensión de salida se puede modelar mediante la aparición de una tensión diferencial en la entrada que se conoce como **tensión de desviación de entrada**.

Estos efectos se irán transmitiendo hacia la salida y darán lugar a la aparición de una tensión de salida distinta de cero que se conoce como **tensión de desviación de salida**.

Por otro lado, aunque la etapa diferencial presenta una ganancia muy alta [2.8] no es infinita y las demás etapas contribuirán algo a la ganancia total pero el conjunto del operacional nunca tendrá las características ideales. Por tanto, **no cabe hablar de ganancia infinita en un operacional real**.

El 741 incorpora dos terminales para el ajuste de la tensión de desviación de entrada. En la figura 2.11 se muestra un potenciómetro de ajuste que permitiría al usuario corregir las diferencias entre las dos ramas y llegar a hacer nula la tensión de salida del operacional cuando sus dos entradas están situadas a masa. Si bien puede hacerse, los cambios introducidos pueden afectar negativamente a otras características dinámicas del operacional.

Finalmente, las entradas "ven" una alta impedancia entre ellas y también ven una impedancia elevada entre cada una de ellas y masa pero, en ambos casos, esa impedancia elevada no resulta infinita por lo que **tampoco es cierta la impedancia de entrada infinita** que se suponía en el operacional ideal.

La etapa intermedia del 741

Su misión es la de adaptar los niveles de tensión ya que el operacional está acoplado en continua y la de realizar la compensación en frecuencia. Su principal efecto se muestra en el comportamiento dinámico del operacional ya que define en bastante medida su ancho de banda mediante el condensador C_1.

La etapa de salida del 741

En la figura 2.12 se muestra el circuito de la etapa de salida del 741 que resulta ser una clásica etapa en contrafase y que incluye las protecciones de sobrecorriente.

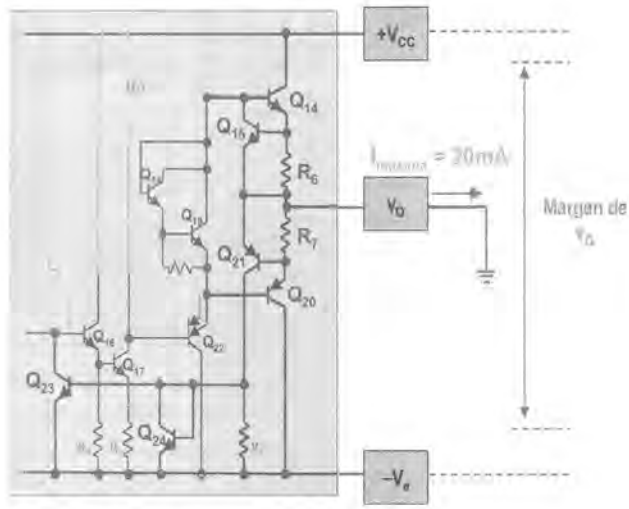


Figura 2.12. Etapa de salida del 741.

La presencia de resistencias en serie con la salida y la propia impedancia de los transistores hacen que **no sea cierta la suposición de impedancia de salida nula**. Por otro lado, los transistores de salida (Q_{14} y Q_{20}) soportarán siempre una caída de tensión —incluso en saturación— que puede llegar a ser pequeña pero nunca nula por lo que **la tensión de salida nunca llegará a alcanzar los valores extremos de la alimentación**. La máxima variación posible en la salida del operacional es la *excursión de la tensión de salida* y suele ir desde un par de voltios por encima de $-V_{ee}$ hasta uno o dos voltios por debajo de $+V_{cc}$. Algunos operacio-

nales que trabajan con tensiones de alimentación muy bajas no pueden permitirse el lujo de perder un margen de salida de más de tres voltios y permiten tensiones de salida mucho más cercanas a las alimentaciones; son los amplificadores ***rail-to-rail***[1] (que se podría traducir como "de extremo a extremo") que prácticamente tienen una excursión de la tensión de salida de $V_{cc}+V_{ee}$.

Aunque las explicaciones hayan sido cualitativas, se ha podido comprobar que el comportamiento del operacional 741 dista de ser el de un dispositivo ideal en el sentido que se ha expuesto aquí; este comportamiento —en mayor o menor medida— puede ser extendido a cualquiera de los operacionales comerciales con los que nos podamos encontrar en el futuro.

Del análisis detallado del comportamiento del circuito interno del 741 se pueden llegar a extraer todos sus parámetros característicos y se puede justificar completamente su comportamiento real [2.14]; sin embargo, en este texto sólo haremos hincapié en el estudio de las características externas sin efectuar la deducción matemática desde el circuito interno ya que, en general, lo que nos va a interesar es el comportamiento del dispositivo como elemento de circuito y cómo influye en el diseño de los bloques y sistemas utilizados en instrumentación.

La caracterización real de un sistema se realiza mediante la determinación de su comportamiento estático y de su comportamiento dinámico según se comentó en el primer capítulo. Siguiendo ese criterio, estudiaremos los parámetros estáticos característicos (parámetros en continua) y luego analizaremos los parámetros dinámicos (parámetros de alterna) que caracterizan al operacional.

2.2.3.1 Características estáticas de los amplificadores operacionales

El comportamiento del operacional ante entradas estáticas o con un cambio lento en el tiempo viene definido por una serie de parámetros que estudiaremos a continuación.

Ganancia diferencial o ganancia en lazo abierto (open loop gain)

La ganancia diferencial resulta no ser infinita aunque su valor puede ser muy elevado. Dependen del operacional pero son frecuentes ganancias diferenciales A_d cercanas o superiores a los 100 dB. Si consideramos un operacional con una ganancia diferencial A_d, la tensión de salida no será exactamente la que se indicaba en la tabla 2.1 para las diversas configuraciones.

El primer efecto que se produce afecta a la conclusión de que $v_d = 0$ con amplificadores realimentados, ya que con A_d finita, como $v_o = A_d v_d$, si la salida no es nula resulta que $v_d \neq 0$. Consideremos el caso de una configuración no inversora como la de la figura 2.13 en la que el operacional resulta ideal en todo excepto en que su ganancia diferencial es finita.

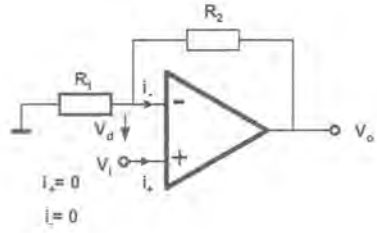


Figura 2.13. Amplificador operacional con ganancia diferencial finita en configuración no inversora.

[1] El término *rail-to-rail* se aplica en dos situaciones diferentes: para definir un amplificador operacional que es capaz de llegar prácticamente a alcanzar en la salida los valores de la alimentación (*rail-to-rail* de salida) y para definir un amplificador operacional que es capaz de admitir (y procesar correctamente) valores de tensión muy próximos a los de alimentación en la entrada (*rail-to-rail* de entrada). En muchos casos, presentan los dos comportamientos, teniendo un amplificador operacional calificado como *input-output rail-to-rail*.

Vamos a calcular la tensión diferencial que se produce en la entrada del operacional. Si a la salida hay v_o, considerando que la corriente por la entrada inversora es nula, el divisor formado por R_1 y R_2 establece una tensión en esa entrada de valor:

$$v_- = \frac{R_1}{R_1 + R_2} v_o \tag{2.9}$$

con lo que la tensión diferencial $v_d = v_+ - v_-$ será:

$$v_d = v_i - \frac{R_1}{R_1 + R_2} v_o \tag{2.10}$$

Como $v_o = A_d v_d$, despejando $v_d = v_o / A_d$, sustituyendo en (2.10) y operando un poco se obtiene que:

$$v_o = \frac{A_d}{1 + A_d \dfrac{R_1}{R_1 + R_2}} v_i \tag{2.11}$$

Expresión que proporciona la tensión de salida de un amplificador no inversor cuando la ganancia diferencial del operacional o ganancia en lazo abierto (*open loop gain*) A_d no es infinita. La expresión (2.11) es similar a la que se presentaba en la tabla 2.1 ya que para $A_d \to \infty$, son iguales.

De una forma paralela a la seguida hasta aquí, pero con la intención de hacer más general el análisis, se considera la fracción de la tensión de salida que aparece en la entrada como βv_o (en el caso anterior, $\beta = R_1/(R_1 + R_2)$) y suponiendo una realimentación negativa o, lo que es lo mismo, que esa fracción de la tensión de salida se resta de la tensión de entrada, se tiene que:

$$v_d = v_i - \beta v_o \tag{2.12}$$

con lo que, en general, tenemos que:

$$v_o = \frac{A_d}{1 + \beta A_d} v_i \tag{2.13}$$

lo que resulta en una expresión general de cualquier sistema realimentado en el que A_d sea la ganancia en lazo abierto y β el factor de realimentación según se indica en la figura 2.14.

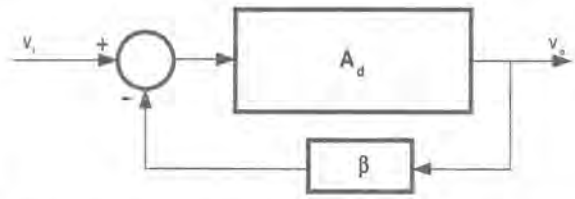


Figura 2.14. Diagrama general de un sistema realimentado con ganancia A_d.

Con ganancia diferencial muy elevada ($A_d \to \infty$), la expresión (2.13) se transforma en:

$$v_o \cong \frac{1}{\beta} v_i \tag{2.14}$$

La pregunta que nos podemos hacer es: ¿qué validez tiene la aproximación (2.14)? Y, en la misma línea: ¿son válidas las expresiones de la tabla 2.1? Desde el punto de vista únicamente de la ganancia diferencial finita resultan bastante válidas ya que la mayoría de los operacionales tienen valores muy altos de A_d.

EJERCICIO 2.1

Consideremos un amplificador diferencial de ganancia teórica 10 como el de la figura 2.15 construido con el 741 que tiene una ganancia en lazo abierto de 106 dB. Considerando que es ideal en todo excepto en ese parámetro, se pide calcular la ganancia de tensión y el error producido con la suposición ideal de ganancia infinita para una entrada diferencial de 1,00000 V. ¿Tiene importancia suponer ganancia infinita en este caso?

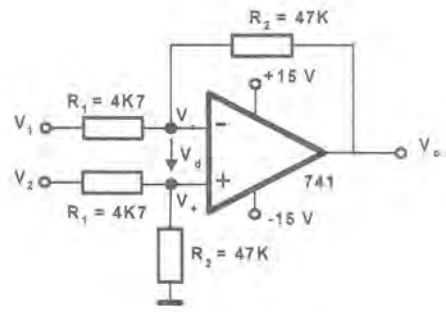


Figura 2.15

SOLUCIÓN:

Según la figura, $R_1 = 4K7$ y $R_2 = 47K$. Vamos a calcular la tensión diferencial que tendrá como causas las dos tensiones de entrada v_1 y v_2 y la fracción de la tensión de salida que aparezca en la entrada; así pues, la tensión en la entrada no inversora será:

$$v_+ = \frac{R_2}{R_1 + R_2} v_2$$

En la entrada inversora, la tensión se deberá a la tensión de entrada v_1 y a la fracción de la tensión de salida que aparezca en la entrada; por superposición:

$$v_- = \frac{R_2}{R_1 + R_2} v_1 + \beta v_o \qquad \text{o} \qquad v_- = \frac{R_2}{R_1 + R_2} v_1 + \frac{R_1}{R_1 + R_2} v_o$$

La tensión diferencial será $v_d = v_+ - v_-$ y como $v_d = v_o/A_d$, tenemos que:

$$\frac{v_o}{A_d} = \frac{R_2}{R_1 + R_2}(v_2 - v_1) - \beta v_o \Rightarrow v_o = \frac{A_d}{1 + \beta A_d} \frac{R_2}{R_1 + R_2}(v_2 - v_1)$$

expresión que con ganancia infinita se convierte en la mostrada en la tabla 2.1. Como $A_d = 106$ dB, es decir, $A_d = 200000$, tenemos que $v_o = 9{,}999945(v_2 - v_1)$; para una entrada diferencial de 1,00000 V, el error producido en la tensión de salida será:

$$\varepsilon = \frac{|10 - 9{,}999945|}{10} = 0{,}000055 \qquad 5{,}5 \cdot 10^{-6} \quad : \quad 5{,}5 \cdot 10^{-4}\,\%$$

Resulta un error relativo muy pequeño para la mayoría de las aplicaciones pero su importancia depende de la precisión buscada.

No obstante, veremos que hay otras no-idealidades cuyas influencias pueden hacerse notar bastante más.

La ganancia diferencial en lazo abierto no tiene un comportamiento constante en función de la frecuencia, sino que va cayendo progresivamente a medida que sube la frecuencia de trabajo.

Resistencias de entrada (input resistances)

En una primera aproximación hay que considerar que la impedancia de entrada no es infinita y que puede aproximarse bastante bien por una resistencia finita, R_{id}, de elevado valor situada entre los terminales de las entradas inversora y no inversora (figura 2.16a).

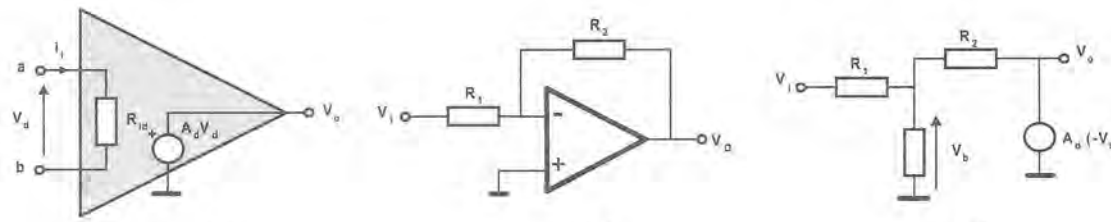


Figura 2.16. Resistencia de entrada de un operacional: (a) modelo básico; (b) circuito inversor; (c) circuito para el cálculo de su efecto sobre el funcionamiento del circuito inversor.

En la figura 2.16b se muestra un circuito amplificador inversor y en la figura 2.16c, el circuito equivalente incluyendo la citada resistencia. La inclusión de R_{id} da lugar a la circulación de corriente a su través, apareciendo la correspondiente tensión diferencial.

Si la ganancia diferencial fuese infinita, el efecto de la resistencia de entrada sería nulo, puesto que la tensión diferencial tendría que ser nula con lo que no circularía corriente por las entradas. Pero con A_d finita, tenemos que:

$$v_d = -v_i \frac{R_2 // R_{id}}{R_1 + R_2 // R_{id}} + v_o \frac{R_1 // R_{id}}{R_2 + R_1 // R_{id}}$$

Teniendo en cuenta que $v_d = v_o / A_d$ y operando un poco se obtiene:

$$v_o = -v_i \frac{A_d}{\frac{R_1}{R_2}(1 + A_d) + 1 + \frac{R_1}{R_{id}}} \qquad (2.15)$$

Como antes comentábamos, si A_d fuese infinita, la expresión obtenida no dependería de R_{id}. Por otro lado, valores bajos de la resistencia de entrada, incrementan el valor del último término del denominador y contribuyen a hacer menos ideal el circuito.

Los valores de resistencia de entrada que se pueden esperar de un operacional real son, en general, bastante elevados y van desde varias centenas de kΩ en el caso de los operacionales con entrada bipolar sencilla hasta 10^6 MΩ o más en el caso de operacionales con entrada FET.

EJERCICIO 2.2

Considerando que la resistencia diferencial del operacional 741 tiene un valor típico de 2 MΩ y un valor mínimo de 300 kΩ se desea calcular el error producido en la salida de un amplificador inversor de ganancia –10 con R_1 = 10K y R_2 = 100K construido usando el 741 cuando a la entrada hay una tensión de 0,10000 V. Considérese A_d = 106 dB.

SOLUCIÓN:

El circuito a considerar es el de la figura 2.16b con los valores indicados de R_1 y R_2; cuando se hace cualquier cálculo en que los valores tienen un margen de variación, siempre hay que considerar el peor caso. Desde el punto de vista de la resistencia de entrada, el peor caso es cuando su valor es más bajo, así que supondremos que $R_{id} = 300K$. En ese caso, la tensión de salida vale, aplicando la expresión (2.15):

$$v_o = -9{,}99943 v_i$$

con lo que la tensión de salida sería $v_o = -0{,}99943$ V, lo que representa un error del 0,06%.

Como en cualquier caso que se pueda presentar, la importancia del error depende de la aplicación y, para la mayoría de los circuitos un error menor del 1% suele ser aceptable. Pero en otras ocasiones resultará excesivo y obligará a elegir otro operacional con mayor resistencia de entrada.

En una segunda aproximación, también deben considerarse las resistencias que aparecen entre las entradas y la masa de referencia del circuito. Estas resistencias son de mayor valor en general que la resistencia diferencial aunque del mismo orden de magnitud. En la figura 2.17a se muestra el circuito con estas resistencias y en la figura 2.17b, el equivalente eléctrico correspondiente a un amplificador diferencial.

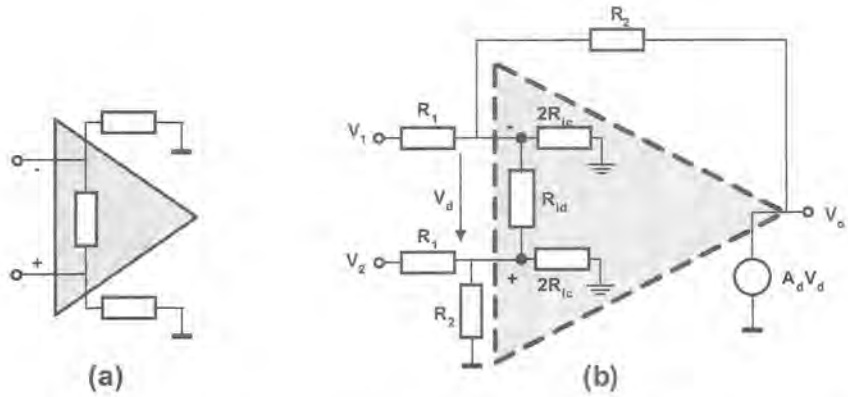


Figura 2.17. (a) Modelo completo de la resistencia de entrada de un operacional y, (b) circuito para el cálculo de su efecto sobre el funcionamiento.

Si ambas son iguales —puede suponerse así en la práctica totalidad de los casos— su efecto puede cuantificarse asumiendo una resistencia de entrada más baja e igual al paralelo de la resistencia de entrada diferencial con las dos resistencias de entrada que conectan las entradas a masa. Así, se puede considerar $R_i = R_{id}//R_{ic}$. Los fabricantes suelen proporcionar este valor como resistencia de entrada y no suelen dar los valores de las dos componentes resistivas por separado, así que no solemos tener como dato ni R_{id} ni $2R_{ic}$ [2.19].

Resistencia de salida (output resistance)

La resistencia de salida del operacional posee, en general, bajos valores puesto que la salida está formada por una etapa en colector común con buenas características en este aspecto; para amplificadores de propósito general, la resistencia de salida oscila entre 40 y 100 Ω mientras que en amplificadores operacionales de potencia es posible encontrar valores bastante más bajos. En la figura 2.18a se muestra un operacional incluyendo la resistencia de salida, en la figura 2.18b un amplificador inversor cargado con una resistencia R_L y

en la figura 2.18c, el equivalente eléctrico para el cálculo del efecto de la resistencia de salida sobre la tensión de salida del circuito.

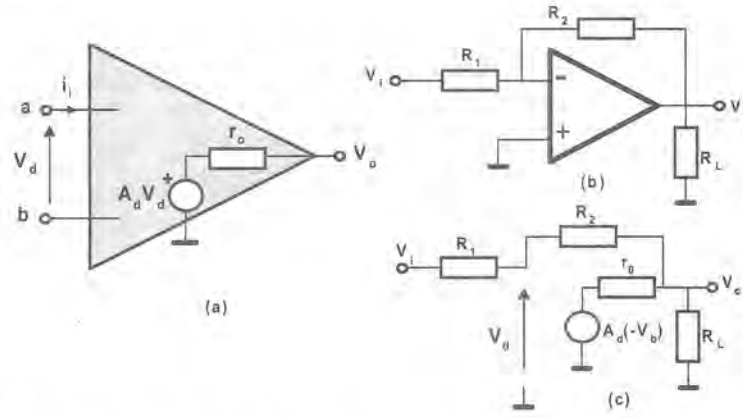


Figura 2.18. Resistencia de salida de un operacional: (a) modelo; (b) circuito con carga; (c) circuito para calcular el efecto sobre la salida.

En el circuito equivalente de la figura 2.18c, se puede observar que la presencia de la resistencia de salida no sólo pone de manifiesto un efecto de carga con pérdida de tensión sobre R_L, sino que evita que la fuente de tensión de salida cortocircuite totalmente la resistencia de carga, haciendo que la resistencia de realimentación R_2 quede en serie con la resistencia formada por el paralelo de R_L y la propia resistencia de salida r_o. Así, v_d vale:

$$v_d = -\frac{R_2}{R_1+R_2}v_i - \frac{R_L}{R_1+R_L}v_o \tag{2.16}$$

Teniendo en cuenta que $v_d = v_o/A_d$ y operando, se tiene que:

$$v_o = \frac{-A_d\dfrac{R_2}{R_1+R_2}+\dfrac{r_o}{R_1+R_2}}{1+A_d\dfrac{R_1}{R_1+R_2}+\dfrac{r_o}{R_1+R_2}+\dfrac{r_o}{R_L}}v_i \tag{2.17}$$

expresión que es igual a la del amplificador inversor con operacional ideal cuando r_o es cero o cuando la ganancia en bucle abierto es infinita, disminuyendo su efecto en la medida que R_L crece.

La importancia de la resistencia de salida no es demasiado alta en cuanto al error que introduce en la tensión de salida pero sí que tiene un papel importante cuando se desea desarrollar una potencia importante sobre la carga como en el caso de los amplificadores de audio. Si es este el caso, hay que seleccionar amplificadores operacionales de potencia capaces de producir algunas decenas de vatios sobre cargas de 4 u 8 Ω.

EJERCICIO 2.3

El amplificador inversor de la figura 2.19 está construido con un 741 cuya ganancia en bucle abierto es de 106 dB y cuya resistencia de salida es 75 Ω. Calcúlese el error introducido en la tensión de salida cuando la entrada es de 0,10000 V. ¿Qué ocurriría si el amplificador estuviese menos realimentado? Considérese el caso en que $R_1 = 1K$ y $R_2 = 50K$.

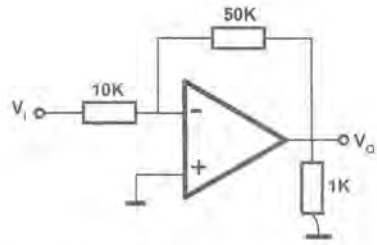


Figura 2.19. Circuito para el ejercicio 2.3.

SOLUCIÓN:

Sustituyendo los valores de R_1, R_2 y la resistencia de carga en la expresión (2.17) obtenemos que la tensión de salida es –0,49998 V cuando debería ser –0,50000 V. El error, en consecuencia es de 0,0032%.

Si tuviésemos un sistema con una realimentación menor (más ganancia), el efecto de la resistencia de salida —como el de otros factores no deseables de la cadena de amplificación directa (sin realimentar)— se verá menos atenuado y aparecerá más claramente en la salida. Así, cuando tengamos un sistema con $R_1 = 1K$ y $R_2 = 50K$, la tensión de salida es de –0,49863 V y el error 0,27%, mayor que antes, según se había previsto.

En cualquiera de estos casos, el efecto sobre el error tiene menos importancia cuantitativa que el debido a la resistencia de entrada para un operacional de tecnología bipolar.

Tensión de desviación de entrada (input offset voltage)

Puede que sea uno de los parámetros más importantes de un operacional a la hora de introducir errores en la tensión de salida. Se debe, como se comentó al principio de este apartado, a la discrepancia de parámetros en las dos ramas de la etapa diferencial de entrada por lo que no es posible predecir cuál será su polaridad en un operacional concreto; sin embargo, la mayoría de los operacionales presentan una distribución aproximadamente normal [2.3]. Así, en la figura 2.20a se indica como una fuente de tensión adicional que puede ser positiva o negativa y que se puede suponer en cualquiera de las dos entradas, optando siempre por la que resulte más cómoda.

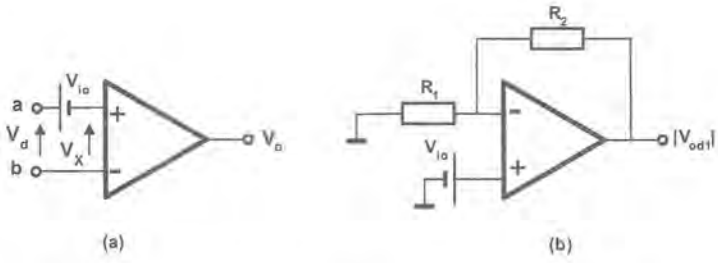


Figura 2.20. Tensión de *offset* de un operacional: (a) *offset* de entrada; (b) *offset* producido en la salida.

Para determinar el efecto sobre la salida, consideremos el circuito de la figura 2.20b (amplificador inversor o no inversor) donde las dos entradas están situadas a masa y sólo se presenta la fuente de tensión correspondiente a la tensión de desviación.

Para esta fuente de tensión, el circuito es un amplificador no inversor y el valor absoluto de la tensión que aparecerá en la salida será:

$$|v_o| = |v_{io}|\frac{R_1 + R_2}{R_1} \tag{2.18}$$

que se denomina *tensión de desviación de salida* y que tiene un signo que depende del signo de la tensión de desviación de entrada.

Este factor no es corregido parcialmente por el efecto de la realimentación ya que está fuera del bucle y para el sistema consta como "una entrada más" que se añade a la señal deseada formando parte del error y que se amplifica tanto o más que la tensión de entrada v_i: en el caso de un amplificador no inversor se amplifica en el mismo factor, mientras que en el caso del amplificador inversor se amplifica un poco más.

Los valores habituales de tensión de desviación de entrada son muy variables de unos amplificadores a otros dependiendo mucho de la tecnología de fabricación. En general, los de entrada FET suelen tener tensiones de desviación muy altas (entre 1 y 20 mV) mientras que los de entrada bipolar presentan valores un orden de magnitud más bajos (entre 0,1 y 2 mV).

Estos valores ya proporcionan una pista de la importancia de la tensión de desviación en un sistema determinado: el error que introducirán será muy importante con tensiones de entrada muy bajas mientras que cuando las tensiones de entrada sean altas, este error tendrá poca importancia.

Si se piensa en un amplificador para instrumentación, en donde los valores de las tensiones de entrada son bajos —a veces comparables con los valores manejados de tensión de desviación— habrá que tener un especial cuidado en la selección de los amplificadores para las etapas de entrada si no queremos tener una señal de salida donde la mayor parte del valor la aporte la tensión de desviación.

EJERCICIO 2.4

Considérese el circuito de la figura 2.21 formado por dos etapas amplificadoras en cascada de ganancia −10 cada una de ellas y construidas tomando como bloque básico el operacional 741 que será considerado como ideal excepto por su tensión de desviación que tiene un valor máximo de 5 mV. Determínese el error introducido en la tensión de salida cuando la entrada vale 50,0 mV.

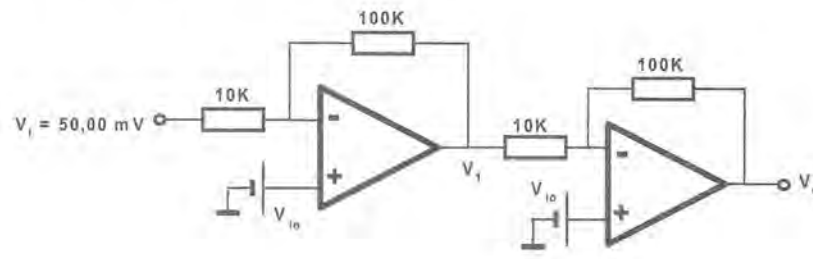


Figura 2.21. Circuito para el ejercicio 2.4.

SOLUCIÓN:

La tensión de salida del primer operacional se debe a la tensión de entrada v_i y a la tensión de desviación v_{io} resultando ser:

$$v_1 = -10v_i \pm \frac{100K + 10K}{10K} v_{io} = -10v_i \pm 11v_{io}$$

El error producido a la salida de la primera etapa resulta ser —considerando $v_{io} = 5$ mV que corresponde al caso más desfavorable— de:

$$\varepsilon = \left| \frac{11v_{io}}{-10v_i} \right| \times 100 = 11{,}0\%$$

¡Un valor considerable! La segunda etapa es similar a la primera pero cuya tensión de entrada sea la de salida de la primera etapa, en consecuencia, la tensión de salida final será:

$$v_o = -10 \times (-10v_i \pm 11v_{io}) \pm 11v_{io} = 100v_i \mp 110v_{io} \pm 11v_{io} \qquad (2.19)$$

Como las polaridades de la tensión de desviación no son predecibles no podemos suponer que se cancelen en ninguna medida y nos tendremos que situar en condiciones de peor caso en el que todas se suman; entonces, la expresión (2.19) se convierte en:

$$v_o = 100v_i \pm 121v_{io}$$

de donde, el error total será del 12,1%. Como se puede observar, la máxima contribución al error aparece en la etapa de entrada que trabaja con menores niveles de tensión mientras que la etapa de salida contribuye en mucha menor medida.

En el ámbito de la instrumentación electrónica es frecuente encontrarnos con valores muy bajos de tensiones que deben ser amplificadas y el panorama que ha dibujado el anterior ejemplo no parece demasiado alentador y realmente no lo es. ¿Qué hacer en estas circunstancias?

La respuesta dependerá de cada caso pero siempre se podrá hacer algo:

a) Si **la tensión que queremos amplificar es alterna**, el problema es más sencillo de resolver: basta con eliminar la tensión continua debida a la desviación mediante un condensador teniendo cuidado que no se sature ninguna etapa.

b) Si **la tensión que queremos amplificar es continua**, el problema es más grave y podemos optar por la compensación de la tensión de desviación o por la selección de un operacional de mejores características respecto a este parámetro. En la figura 2.22 se muestran dos posibles soluciones para la compensación de la tensión de desviación ya sea utilizando los terminales que al efecto tienen algunos operacionales ya sea añadiendo externamente una tensión igual a la desviación pero de sentido contrario.

La solución propuesta en la figura 2.22a no siempre es factible ya que algunos operacionales carecen de terminales de ajuste; aun en caso de tenerlos hay que actuar con bastante precaución ya que se está modificando el comportamiento del circuito interno del operacional y esto puede producir efectos negativos sobre las especificaciones de partida, como en algunos operacionales en que se produce un empeoramiento de la respuesta dinámica.

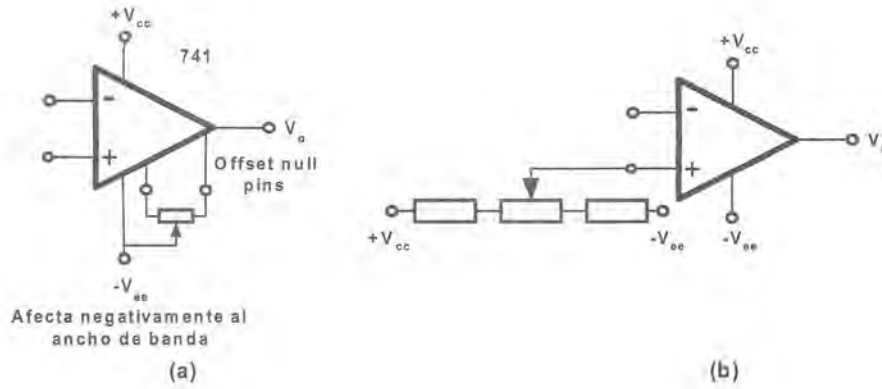


Figura 2.22. Ajuste de la tensión de *offset* mediante el uso de los pines a tal efecto (a) o usando un circuito externo (b).

En cualquier caso se puede llevar a cabo un ajuste mediante el potenciómetro de la figura 2.22b que permitirá compensar la tensión de desviación al añadir un valor igual pero de signo contrario. No obstante, a la hora de efectuar un diseño para fabricar en serie hay que valorar la posibilidad de efectuar ajustes en el circuito porque lo encarece sustancialmente al precisar mano de obra que ejecute el protocolo de ajuste correspondiente.

En esta línea, suele ser mejor opción buscar un operacional que tenga mejores características (aunque su coste resulte más elevado) y que introduzca un error asumible, haciendo innecesario el ajuste; a él sólo se acudirá como última alternativa.

EJERCICIO 2.5

Se pretende amplificar una tensión senoidal de 10 mV de amplitud y una frecuencia de 100 Hz para obtener una salida de 10 V de amplitud usando operacionales rail-to-rail ideales en todo excepto en la tensión de desviación que es de 5 mV y alimentados simétricamente con 12 V.

SOLUCIÓN:

La tentación inmediata sería la de efectuar el paso de amplificación en una etapa. Esto introduce otros riesgos que afectan a la respuesta dinámica y que se estudiarán más adelante, pero aquí sólo haremos hincapié en el efecto de la tensión de desviación. Si hacemos la amplificación en una sola etapa, la ganancia necesaria sería de 1000 (60 dB) para lo que diseñaríamos una etapa inversora con, por ejemplo, R1 = 500 Ω y R2 = 500K.

La salida sería:

$$v_o = -10 \operatorname{sen}(2\pi 100t) \pm \frac{500 + 500K}{500} v_{io} = -10 \operatorname{sen}(2\pi 100t) \pm 5{,}005 \text{ V}$$

Expresión que puede llegar a valer +15 V o –15 V (según el signo de v_{io}) y que, con el operacional alimentado a ±12 V, ocasionaría su saturación y el consiguiente recorte de la onda de tensión de salida. Aunque se situase un condensador a la salida que eliminase el nivel de continua, la distorsión resultaría inevitable (figura 2.23b).

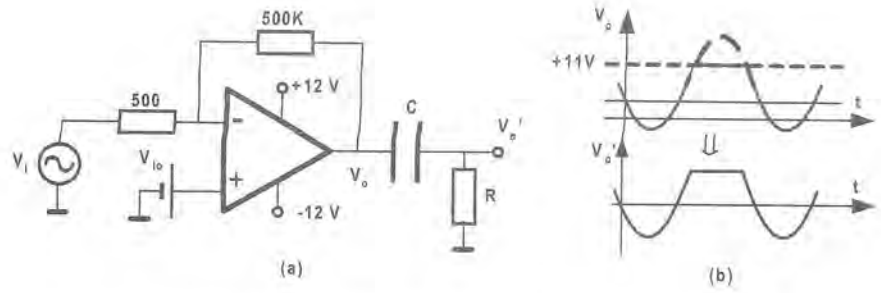


Figura 2.23. (a) Circuito posible para resolver el ejercicio 2.5; (b) formas de onda con este circuito.

Para resolver el problema se puede optar por realizar la amplificación en dos etapas acopladas en alterna con una ganancia importante (–100) en la primera tal y como se muestra en la figura 2.24.

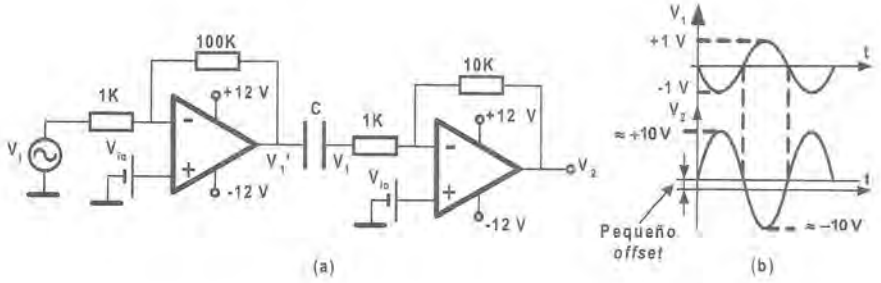


Figura 2.24. Solución al ejercicio 2.5: (a) circuito y, (b) formas de onda con un pequeño *offset* final.

A la salida de la primera etapa la tensión que aparece es:

$$v_1' = -1 \text{ sen } (2\pi\, 100t) \pm \frac{100K + 1K}{1K} v_{io} = -1 \text{ sen } (2\pi\, 100t) \pm 0{,}505 \text{ V}$$

El condensador se cargará con el valor de continua dejando pasar la alterna (si su capacidad es suficientemente grande) v_1 que la segunda etapa amplificará con ganancia –10 proporcionando una tensión de:

$$v_o = 10 \text{ sen } (2\pi\, 100t) \pm \frac{10K + 1K}{1K} v_{io} = 10 \text{ sen } (2\pi\, 100t) \pm 0{,}055 \text{ V}$$

La desviación de la salida es muy pequeña y representa un error poco mayor del 0,5% pero si aún se desease eliminarlo completamente, podríamos situar un condensador en serie con la salida para eliminarlo tal y como se había hecho en el circuito de la figura 2.23.

EJERCICIO 2.6

Se dispone de un operacional 741 alimentado simétricamente entre –12 y +12 V que se va a usar para amplificar en una etapa una señal continua procedente de un sensor y que puede tomar cualquier valor entre –100,00 y +100,00 mV hasta llevarla a una escala respectiva entre 1 y –1 V para lo que se elige una etapa

inversora con R_1 = 4K7 y R_2 = 47K. Determínese el error a fondo de escala que se produce en la salida si v_{io} tiene un valor máximo de 5mV y diséñese una red de compensación como la de la figura 2.22b para tratar de reducir el error.

SOLUCIÓN:

En la etapa inversora, la tensión de salida debida a la tensión de desviación viene dada por (2.18) con lo que la tensión de salida causada por la tensión de desviación resulta ser de ±0,055 V y el error provocado por este concepto resulta ser de 5,5%. Resulta, claro está, un valor muy elevado que debería ser corregido para lo que se puede diseñar una red de compensación como la de la figura 2.22b.

Como quiera que la tensión que debe ser compensada debe ser de 5 mV como máximo, la red externa debería ser capaz de proporcionar cualquier valor comprendido entre –5 y 5 mV; si se parte de las tensiones de alimentación a ambos extremos del divisor resistivo (figura 2.25) la tensión en el punto A (v_A) resulta ser:

$$v_A = \frac{R}{2R+P}24 - 12 = -5 \cdot 10^{-3} \text{ V}$$

con lo que la relación entre P y R resulta ser de R = 1200,5P y cualquier pareja de resistencias que cumpla esa relación permitirá dejar en los extremos del potenciómetro tensiones de valor –5 y 5 mV. En la práctica, con resistencias que no tienen un valor exacto sino una tolerancia, v_A y v_B serían algo mayores (en valor absoluto) para permitir un ajuste en el peor caso, obteniendo que R > 1200,5P lo que nos llevaría, por ejemplo a un potenciómetro de 100 Ω y una resistencia algo superior a 120K (se elegiría el valor estándar más próximo teniendo en cuenta que la tolerancia de la resistencia no llegue a dejar el valor por debajo de 120050 Ω.

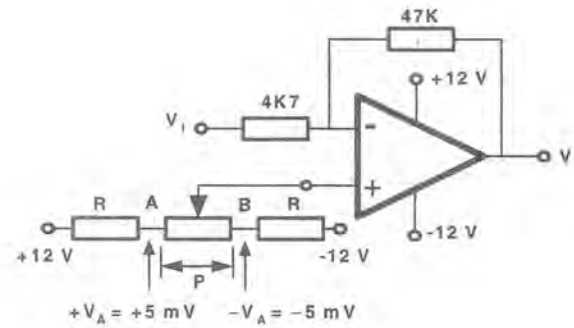


Figura 2.25. Ajuste externo de *offset* para el ejercicio 2.6.

Una vez situado el potenciómetro de ajuste en el circuito el procedimiento para eliminar la tensión de desviación sería:

1. Situar la entrada a 0 V.
2. Medir la tensión de salida.
3. Actuar sobre el potenciómetro hasta conseguir una lectura de 0 V en la salida.

Hay que recordar que el situar puntos de ajuste requiere —siempre— un protocolo de calibrado y que si no se hace así, podremos encontrarnos con el potenciómetro situado en el extremo opuesto por aplicación directa de la Ley de Murphy[2] [2.4], consiguiendo llegar a duplicar el error por este concepto.

El problema de la tensión de desviación es bastante grave ya que produce errores importantes en la tensión de salida pero el problema es aún peor de lo que parece puesto que el método de compensación puede llegar a ser inútil debido a efectos adicionales: hablamos de las derivas de tensión (*voltage drift*).

Este es un efecto que causa un grave problema en los amplificadores operacionales y que ocasiona la variación de los parámetros por dos motivos:

a) Debido a cambios en la temperatura (*thermal voltage drift*) denotado por $\Delta v_{io}/\Delta T$ que provoca alteraciones en el comportamiento de los dispositivos que constituyen internamente el operacional. Las derivas hacen que el valor de la tensión de desviación de entrada no sea constante con la temperatura, sino que variará conforme se modifique la temperatura del operacional de tal manera que el ajuste que se pudiera haber hecho en unas condiciones de temperatura dadas, se convierte en no válido para otras.

b) Debido a cambios a lo largo de la vida del operacional y que se pueden nombrar como inestabilidades en los valores de las tensiones (*long-term input offset voltage stability*) denotado por $\Delta v_{io}/\Delta t$.

Ninguna de las dos es predecible y sólo puede acotarse su valor por lo que, para tenerlas en cuenta sólo quedan dos posibilidades:

1. Calcular su efecto y considerarlo como despreciable si queda dentro de unos márgenes determinados de error en un plazo razonable y con una variación previsible de temperatura.
2. Efectuar una calibración cada vez que se vaya a utilizar el equipo para garantizar que durante un periodo de tiempo se puede utilizar con pequeñas variaciones.

Los valores de las derivas califican en buena medida el comportamiento de los amplificadores operacionales ya que determinan la invariabilidad de la tensión de desviación frente a cambios en las condiciones ambientales y durante periodos largos de uso.

Por ejemplo, un operacional como el 741 presenta un valor máximo de 15 μV/°C lo que significa que la tensión de desviación de entrada puede variar bastante si cambia la temperatura. El valor de las derivas en el tiempo no suele ser un parámetro habitual en las hojas de características de los fabricantes, excepto en los casos de amplificadores para uso en instrumentación, aunque los valores suelen ser siempre pequeños: también a modo de ejemplo, el OPA177F presenta una deriva de 0,3 μV/mes.

Corrientes de polarización (*input bias currents*)

Las corrientes de polarización representan los valores de las corrientes que circulan por las dos entradas del operacional y son necesarias para la puesta en funcionamiento de la etapa diferencial de entrada (ver figura 2.12). Si las dos ramas de la entrada diferencial fuesen exactamente iguales, ambas corrientes también lo serían pero no es posible predecir cuál de las dos va a conducir más y, en consecuencia, sus valores serán similares pero diferentes. En la figura 2.26 se indica el modelo utilizado presentando las dos corrientes i_{B1} e i_{B2} y las resistencias que se ven desde cada una de las entradas, R_{Th-} y R_{Th+}.

[2] La Ley de Murphy, un capitán estadounidense de la base de Edwards, dice "*Si algo puede ir mal, irá mal*". Su aplicación a diversos ámbitos de la técnica ha sido recopilada en muchas frases célebres y divertidas. Su principal impulsor ha sido Arthur Bloch cuya principal referencia se cita aquí.

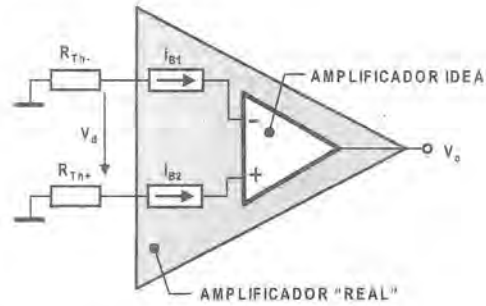


Figura 2.26. Las corrientes de polarización del operacional causan una tensión de *offset* de entrada adicional.

La circulación de esas corrientes origina una tensión diferencial de valor:

$$v_d = R_{Th+} i_{B1} - R_{Th-} i_{B2} \tag{2.20}$$

No es posible predecir el valor de las corrientes, y los fabricantes sólo proporcionan el valor medio de ambas y la diferencia entre ellas sin indicar cuál es la mayor y cuál la menor. Así, cualquier fabricante de operacionales proporciona en sus hojas de características el valor de la corriente de polarización (*input bias current*):

$$I_B = \frac{|i_{B1} + i_{B2}|}{2} \tag{2.21}$$

y el de la corriente de desviación (*input offset current*) que representa la variación prevista entre ambas:

$$I_{io} = |i_{B1} - i_{B2}| \tag{2.22}$$

El valor de la corriente de desviación suele ser en torno a un 25% del valor de la corriente de polarización; ambos son muy variables en función de la tecnología de fabricación del operacional encontrando valores elevados en operacionales de tecnología bipolar (entre 10 y 100 nA), valores bajos con entradas FET (entre 1 y 10 pA) y extraordinariamente bajos en el caso de entradas de tipo MOSFET (menores de 0,001 pA). El caso del operacional que ponemos como referencia, el 741 de tecnología bipolar, tiene valores típicos de corrientes de polarización de 80 nA aunque presenta valores máximos de hasta 500 nA; como valores de la corriente de desviación, su valor típico es de 20 nA aunque puede llegar hasta máximos de 200 nA.

Poniendo la expresión (2.20) en función de los parámetros definidos en (2.21) y (2.22) que son los que encontraremos en las hojas de características de cualquier operacional, tenemos que la tensión de desviación V_{dB}:

$$v_{dB} = R_{Th+}\left(I_B + \frac{I_{io}}{2}\right) - R_{Th-}\left(I_B - \frac{I_{io}}{2}\right)$$

Operando, obtenemos:

$$v_{dB} = I_B\left(R_{Th+} - R_{Th-}\right) + I_{io}\left(R_{Th+} + R_{Th-}\right)/2$$

Como no es predecible si el valor de i_{B1} es mayor que el de i_{B2}, el valor máximo de la tensión de desviación será:

$$v_{dB} = I_B\left|R_{Th+} - R_{Th-}\right| + I_{io}\left(R_{Th+} + R_{Th-}\right)/2 \tag{2.23}$$

Lo que significa que siempre que las resistencias que se vean desde las dos entradas del operacional sean iguales ($R_{Th-} = R_{Th+}$), la corriente de polarización no tendrá efecto sobre la tensión de desviación ni, por ello,

sobre la salida. Por el contrario, la corriente de desviación sí que contribuye a la tensión de desviación en todas las circunstancias haciendo más notoria su participación en la medida en que los valores de las resistencias resulten altos.

La tensión de salida, considerando los efectos de las corrientes de polarización, dependerá de la ganancia de la configuración elegida y, en general, se puede aplicar el mismo proceso que si se tratase de la tensión de desviación de entrada.

En algunas configuraciones la resistencia que se ve desde la entrada inversora y de la no inversora es igual, como en el caso del amplificador diferencial por lo que el efecto de la corriente de polarización se anula según (2.23) al provocar la misma caída de tensión sobre la resistencia. En otras no ocurre así y en estos casos se suele encontrar la recomendación de añadir una resistencia adicional que, situada adecuadamente, no modifique la función del circuito y elimine el problema de la corriente de polarización. En la figura 2.27 se muestran algunos ejemplos de empleo de la resistencia de compensación.

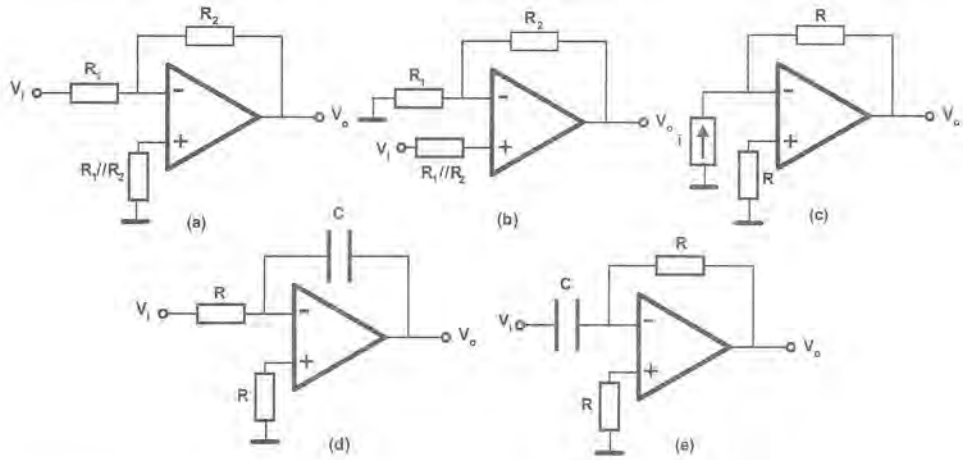


Figura 2.27. Empleo de la resistencia que compensa el efecto de la corriente de polarización sobre diversos circuitos.

La solución de la figura 2.27 presenta algunos problemas relacionados con la generación de ruido sobre la resistencia de compensación por lo que habrá que extremar las precauciones cuando se recurra a este tipo de soluciones.

Además de lo comentado antes, hay que tener en cuenta que cuando las corrientes de polarización y desviación tienen importancia en la tensión de salida, se debe recurrir a amplificadores con bajos valores de I_B e I_{io}. En estos casos, los valores de las corrientes de polarización y de desviación son comparables y, en consecuencia, el efecto de la segunda puede ser mayor que el efecto de la primera.

¡Pero el efecto de la corriente de desviación no puede eliminarse! Entonces, la única opción que queda es reducir el valor de las resistencias involucradas en el circuito por lo que el factor más decisivo resultará ser en todos los casos, el de la corriente de desviación.

EJERCICIO 2.7

Determinar el error que se produce en la salida de un amplificador inversor con v_i = 100,00 mV en el que R_1 = 10K y R_2 = 47K construido con un operacional ideal en todo excepto en la presencia de una corriente

de polarización I_B = 100 nA y una corriente de desviación I_{io} = 25 nA. Estudiar el efecto que tiene añadir una resistencia de compensación en la entrada no inversora.

SOLUCIÓN:

La resistencia que se ve desde la entrada inversora es 10K en paralelo con 47K, es decir, R_{Th-} = 8,2456 kΩ mientras que en la entrada no inversora R_{Th+} = 0 Ω. De la expresión (2.23), se obtiene que la tensión de diferencial producida será de 0,92763 mV.

A la salida aparecerá una tensión debida a las corrientes de polarización de valor:

$$v_{oB} = \frac{R1+R2}{R1} v_d = \frac{10K+47K}{10K} 0{,}92763 \cdot 10^{-3} = 5{,}2875 \text{ mV}$$

Como la tensión de salida debería ser de –470,00 mV, el error que se produce es algo mayor del 1,1%. Para compensar la corriente de polarización situaríamos una resistencia de 8,2456 kΩ en la entrada no inversora (ver figura 2.28a) o una de valor muy próximo como la de valor estándar de 8K2.

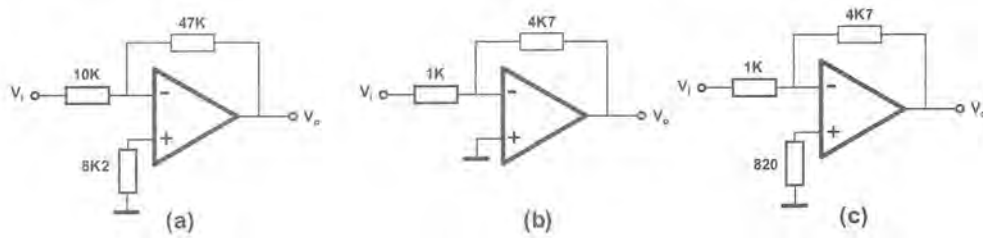


Figura 2.28. Reducción de los efectos de las corrientes de polarización y desviación para el ejercicio 2.7: (a) uso de la resistencia de compensación; (b) reducción de los valores óhmicos de las resistencias; (c) las dos acciones simultáneamente.

Para este caso, tenemos que la tensión diferencial sería v_{dB} = 0,41570 mV y se deberá en su mayor parte a la corriente de desviación. El error producido ahora será un poco más del 0,50%, es decir, sólo la mitad del anterior. Si, por el contrario, construyésemos el circuito con resistencias diez veces menores (figura 2.28b), la reducción del efecto sería mucho mayor, con una tensión de desviación de sólo 0,092763 mV y un error total del 0,11%. Si, además, se compensase el efecto de la corriente de polarización con una resistencia de 820 Ω en la entrada no inversora (figura 2.28c) el error bajaría hasta el 0,050%.

Como quiera que el efecto de las corrientes de polarización se traduce en una tensión diferencial, puede ser considerado igual al de la tensión de desviación de entrada, así que se sumarán ambos efectos y las imperfecciones de la etapa de entrada, sea cual sea su causa, se traducirán en un error con la aparición de una tensión de desviación de salida.

Cabría pues utilizar un método de corrección igual al que se planteaba en la figura 2.22b que compense ambos efectos en el supuesto de no utilizar un operacional con unas prestaciones suficientemente aceptables.

Aun en el caso de una compensación por ajuste, el problema térmico persiste y también en el caso de las corrientes de polarización se producen derivas por las mismas causas que antes se estudiaron:

a) Derivas térmicas en la corriente de desviación (*thermal current drift*) denotado por $\Delta I_{io}/\Delta T$ y que, en el 741, toma un valor típico de 0,5 nA/°C.

b) Derivas con el tiempo (*long-term input offset current stability*) denotado por $\Delta I_{io}/\Delta t$.

EJERCICIO 2.8

Un amplificador inversor de ganancia –10 con R_1 = 10K y R_2 = 100K está montado con un operacional 741 cuyas derivas térmicas resultan ser de 15 μV/°C para la tensión de desviación y 0,5 nA/°C para la corriente de desviación tiene que trabajar con una variación de temperatura amplia, considerándose que puede estar sometido a cualquier valor de temperatura entre 10° y 70°C. Suponiendo que se dispone de un ajuste de tensión de desviación de salida según el esquema de la figura 2.22b, que está realizado a 25°C y que se añade una resistencia de compensación de la corriente de polarización I_B, determínese el error producido cuando la entrada es de 100,0 mV.

SOLUCIÓN:

Supongamos el amplificador ajustado a 25°C con lo que la tensión de desviación de salida será nula para ese valor. Al trabajar a una temperatura diferente, los valores de tensión de desviación de entrada y de corriente de desviación se van a modificar y serán más diferentes cuanto más se aleje la temperatura de la de ajuste, es decir, el máximo error se producirá cuando T = 70°C.

Para ese caso, tenemos que Δv_{io} = 15 μ/°C × 45°C = 675 μV y ΔI_{io} = 0,5 nA/°C × 45°C = 22,5 nA con lo que la tensión diferencial será de v_d = 675 μV + 22,5 nA×10K//100K = 879 μV y la tensión de desviación de la salida se verá amplificada en el factor $(R_1 + R_2)/R_1$ resultando un valor de 9,669 mV lo que supone un error en la tensión de salida de un 1% aproximadamente.

Lógicamente, valores más bajos de la tensión de entrada llevarían a errores que podrían llegar a ser inaceptables y que no pueden ser compensados salvo por ajuste antes de uso. Por este motivo, se reafirma lo que se había comentado antes y es que resulta trascendental una adecuada selección de los operacionales en función de las derivas que presenten por cuanto suponen el principal factor de incertidumbre.

Razón de rechazo del modo común (Common Mode Rejection Ratio)

Los amplificadores operacionales son capaces de producir una ganancia diferencial muy elevada pero, como hemos visto, no infinita. Además, incorporan un segundo término de ganancia que se encarga de multiplicar la *tensión de modo común*. Para comprender este concepto consideremos el caso del amplificador no inversor que se muestra en la figura 2.29a y que está sometido a una entrada v_i.

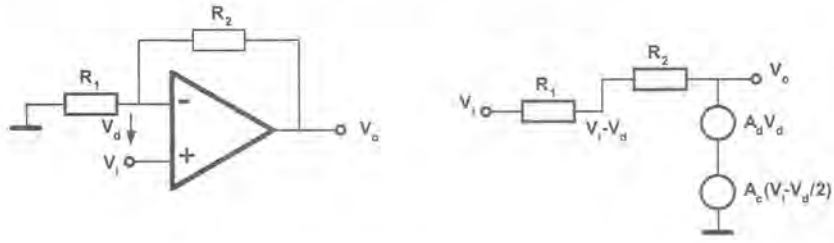


Figura 2.29. Efecto de la ganancia de modo común sobre la configuración no inversora.

Considerando la ganancia diferencial en lazo abierto, la tensión de salida es la determinada por (2.11). Pero, por el hecho de que haya aplicada una tensión en la entrada no inversora, aparecerá una tensión casi igual (sólo se diferencian en v_d) en la entrada inversora. Se denomina tensión de modo común a la media aritmética de la tensión en las entradas del operacional, es decir:

$$v_c = \frac{v_- + v_+}{2} \approx v_i \tag{2.24}$$

Pues bien, el operacional también amplifica este valor en un factor conocido como *ganancia de modo común* A_c con lo que la tensión de salida será la suma de la debida a la ganancia diferencial y la provocada por la ganancia de modo común, es decir:

$$v_o = A_d v_d + A_c v_c \tag{2.25}$$

contribuyendo al error total del operacional. En general, la ganancia de modo común es muy baja en los operacionales reales y suele venir definida por el parámetro conocido como **razón de rechazo de modo común** o por sus siglas en inglés (CMRR) que se define como el cociente entre la ganancia diferencial y la ganancia de modo común y suele expresarse en decibelios:

$$\text{CMRR} = 20\log\frac{A_d}{A_c}\ \text{(dB)} \tag{2.26}$$

Para el 741, el valor es de 90 dB, es decir, el valor de la ganancia diferencial es muy alto en comparación con el de la ganancia de modo común.

Para calcular el efecto sobre la tensión de salida se considera el modelo de la figura 2.29b en la que se muestran las dos fuentes dependientes que aportan la ganancia diferencial y de modo común. La expresión de la tensión de salida resulta ser con ese modelo:

$$v_o = \frac{A_d + A_c/2}{1 + \left(A_d + \frac{A_c}{2}\right)\frac{R_1}{R_1 + R_2}} v_i \tag{2.27}$$

Como habitualmente $A_d >> A_c$, la expresión puede aproximarse por:

$$v_o = \frac{A_d + A_c/2}{1 + A_d \frac{R_1}{R_1 + R_2}} v_i \tag{2.28}$$

Aunque pueda parecer que el efecto que se va a producir va a ser muy pequeño, hay que tener en cuenta que en muchas ocasiones tenemos señales que deseamos amplificar y que presentan un valor de tensión de modo común muy elevado como puede ocurrir en los amplificadores diferenciales. Sea el amplificador diferencial de la figura 2.30 en el que se han incluido las fuentes en la salida que modelan la ganancia diferencial y la de modo común.

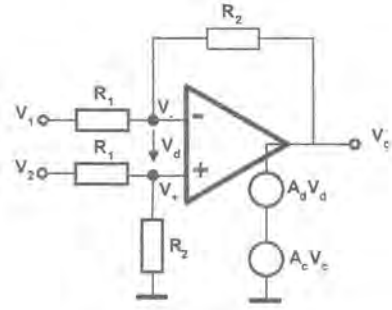


Figura 2.30. Cálculo del efecto de la ganancia de modo común en un amplificador diferencial.

Las tensiones en las entradas no inversora e inversora son, respectivamente:

$$v_+ = \frac{R_2}{R_1 + R_2} v_2 \quad v_- = v_+ - v_d = \frac{R_2}{R_1 + R_2} v_2 - v_d$$

Suponiendo la corriente nula en la entrada inversora, la corriente que circula por R_2 y por R_1 son iguales, es decir:

$$\frac{v_o - \left(\frac{R_2}{R_1 + R_2} v_2 - v_d \right)}{R_2} = \frac{\frac{R_2}{R_1 + R_2} v_2 - v_d - v_1}{R_1} \tag{2.29}$$

En la salida del operacional, $v_o = A_d v_d + A_c v_c$ y la tensión de modo común resulta ser:

$$v_c = \frac{v_+ + v_-}{2} = \frac{R_2}{R_1 + R_2} v_2 - \frac{v_d}{2}$$

Despejando v_d en función de v_o y v_2 y sustituyendo en (2.29) se obtiene que la tensión de salida en función de las tensiones de entrada resulta ser:

$$v_o = \frac{R_2}{R_1} \frac{\left(A_d - \frac{A_c}{2} \right)}{A_d - \frac{A_c}{2} + \frac{R_1 + R_2}{R_1}} (v_2 - v_1) + \frac{R_2}{R_1} \frac{A_c}{A_d - \frac{A_c}{2} + \frac{R_1 + R_2}{R_1}} v_2 \tag{2.30}$$

apareciendo un primer término debido a la tensión diferencial y que se puede aproximar a la ganancia ideal para ganancias del circuito bajas y un segundo término que se debe al modo común y que es menor en la medida en que v_2 sea baja. La expresión (2.30) se puede aproximar sin demasiados problemas considerando que $A_d >> -A_c/2 + (R_1 + R_2)/R_1$ lo que es cierto en la mayoría de los casos, obteniendo una expresión bastante útil:

$$v_o = \frac{R_2}{R_1} (v_2 - v_1) + \frac{R_2}{R_1} \frac{v_2}{CMRR} \tag{2.31}$$

Esta expresión nos indica que el amplificador diferencial, teniendo en cuenta la CMRR, funciona como el ideal pero añadiendo un término que depende directamente del valor de v_2 lo que implica que el error crecerá cuando tengamos tensiones elevadas respecto de masa.

EJERCICIO 2.9

Se considera un amplificador diferencial de ganancia 100 (R_1 = 1K y R_2 = 100K) construido con un operacional 741 que tiene una CMRR de 90 dB y una ganancia diferencial en lazo abierto de 106 dB. Calcúlese el error en la tensión de salida cuando v_1 = 15,000 V y v_2 = 15,010 V. ¿Cuánto valdría el error si las señales v_1 y v_2 tuviesen el mismo valor diferencial pero un modo común de sólo 1,0050 V?

SOLUCIÓN:

Considerando la expresión (2.31), la tensión de salida sería:

$$v_o = 100 \quad (15{,}010 - 15{,}000) + 100 \frac{15{,}010}{31623} = 1{,}0000 + 0{,}047465 \text{ V}$$

lo que supone una tensión de 1,0000 V (la que debería ser en condiciones ideales) más una tensión correspondiente al modo común que implica un error de un 4,7% que resulta un valor considerable. Obsérvese que la información está contenida en la diferencia mientras que el valor de modo común no es deseado. Si hubiésemos aplicado la expresión "exacta" (2.30), obtendríamos un resultado parecido: con $A_d = 200000$ y CMRR = 31623, tenemos que $A_c = 6{,}3245$. Sustituyendo los valores en la expresión (2.30) tendremos que v_o = 0,99950 + 0,047442 V, valores casi iguales a los obtenidos con la expresión aproximada.

Para el caso en que el modo común sea menor, la tensión de salida valdría:

$$v_o = 100\ \ (1{,}0100 - 1{,}0000) + 100\frac{1{,}0100}{31623} = 1{,}0000 + 0{,}0031938\ \text{V}$$

que proporciona un error poco mayor de 0,3%. El efecto de la tensión de modo común a la entrada resulta patente y da lugar a problemas serios en los sistemas en que tienen que trabajar en estas condiciones. Desafortunadamente, no suele haber amplificadores operacionales con mejores prestaciones en este aspecto.

EJERCICIO 2.10

Determínese la tensión de salida del circuito de la figura 2.31 cuando la tensión diferencial de entrada es v_d = 10,000 mV y la tensión de modo común v_c es una senoide de 5 V de amplitud y 50 Hz de frecuencia producida por una interferencia externa de la red. Considere que el operacional utilizado tiene una CMRR de 90 dB. Considerando que la interferencia es un ruido, ¿cuál es la relación S/N a la salida del circuito? Compárela con la relación S/N a la entrada para valorar la función del amplificador diferencial.

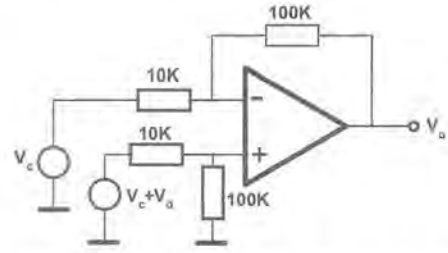


Figura 2.31. Circuito del ejercicio 2.10.

SOLUCIÓN:

La señal de salida será, aplicando la expresión aproximada (2.31):

$$V_o = \frac{100K}{10K}10{,}000\cdot10^{-3} + \frac{100K}{10K}\frac{5\ \text{sen}\ (2\pi 50t)}{31623} = 0{,}10000 + 1{,}5811\cdot10^{-3}\text{sen}\ (2\pi 50t)\ \text{V}$$

El modo común prácticamente ha desaparecido de la señal luego esto nos indica que una de las principales aplicaciones del amplificador diferencial es la eliminación de ruidos de modo común que son fuertemente amortiguados por la propia topología.

Si calculamos la relación S/N a la entrada y a la salida, tendremos que, a la entrada del sistema, la relación es muy baja, en concreto $(S/N)_i = -51{,}0$ dB mientras que a la salida tendremos $(S/N)_o = +39{,}0$ dB.

En el ejemplo anterior aparece una imagen clara del amplificador diferencial como sistema capaz de reducir el nivel de ruido en modo común —habitual en muchos casos— de una señal. Sin embargo, pensar que

esto se consigue en todos los casos resulta inexacto puesto que la ganancia diferencial y la ganancia de modo común tienen distinto comportamiento en función de la frecuencia. Este aspecto se concretará más adelante.

Razón de rechazo de la alimentación (Power Supply Rejection Ratio)

La tensión de alimentación del operacional afecta a todo el circuito y aunque la etapa de entrada sea bastante inmune a sus efectos gracias a la presencia de una fuente de corriente para su polarización, sí que influye en la etapa de salida y en la etapa intermedia. El efecto no suele ser demasiado grande pero sí que debe tenerse en cuenta. La forma de considerarlo es mediante el error que introduce en la tensión de salida Δv_o y que permite definir la **razón de rechazo de la tensión de alimentación** o PSRR como:

$$\text{PSRR} = 20\log\frac{\Delta V_{alim}}{\Delta v_o} \tag{2.32}$$

donde $V_{alim} = V_{cc} - V_{ee}$. El valor de PSRR suele variar de un tipo de operacional a otro entre 60 y 120 dB y representa sólo el efecto que el cambio de la alimentación tiene en condiciones estáticas o de baja frecuencia, es decir, variaciones lentas de la tensión de alimentación. Si las variaciones fuesen rápidas, el cambio no quedaría reflejado correctamente por este parámetro ya que ese tipo de señales afecta muy directamente a la salida del operacional. Este tipo de efecto se estudiará más en detalle en el apartado relacionado con los ruidos.

El valor de PSRR resulta muy importante en sistemas portátiles de instrumentación que estén alimentados por baterías o pilas en los que la tensión de alimentación irá variando en función de su nivel de carga. Si se procede a la alimentación directa desde la batería, nos podremos encontrar con un error variable en función del tiempo de uso del equipo (y del nivel de carga) según se indica en la figura 2.32a. Para resolver el problema, se puede situar un regulador de tensión en la alimentación que amplíe el margen de uso, manteniendo el nivel de tensión constante en V_{cc} (figura 2.32b). Lógicamente, el precio a pagar es una pérdida adicional de potencia por el rendimiento del regulador que siempre será menor que 1.

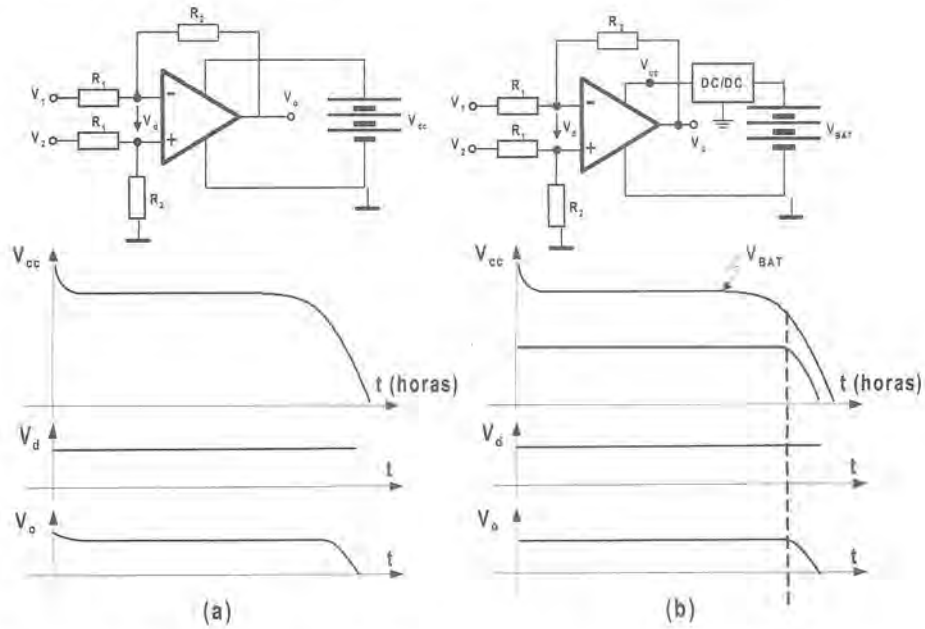


Figura 2.32. Efecto de la variación de la alimentación sobre la tensión de salida en un sistema alimentado desde una batería: (a) sin regulador de tensión; (b) con regulador de tensión.

2.2.3.2 Características dinámicas de los amplificadores operacionales

En este apartado, seguiremos con la caracterización del operacional real, es decir, su comportamiento cuando está sometido a señales que varían su valor a una determinada frecuencia o que producen transiciones entre dos niveles diferentes. El amplificador operacional real no es un componente muy rápido y pondrá una serie de limitaciones cuando deba producir cambios rápidos en su salida.

Ancho de banda (bandwidth)

La ganancia A_d de un operacional no presenta las excelentes características que se han comentado a todas las frecuencias. En general, su valor es muy alto (alrededor de 100 dB) a frecuencias bajas pero va bajando a medida que la frecuencia sube. En la figura 2.33 se puede observar la representación de la ganancia en lazo abierto de un operacional 741 en que se ve que el valor de 106 dB sólo se tiene para $f < 4$ Hz mientras que, de ahí en adelante, cae con una pendiente de –20 dB/déc hasta llegar a 0 dB a una frecuencia de 1,5 MHz aproximadamente.

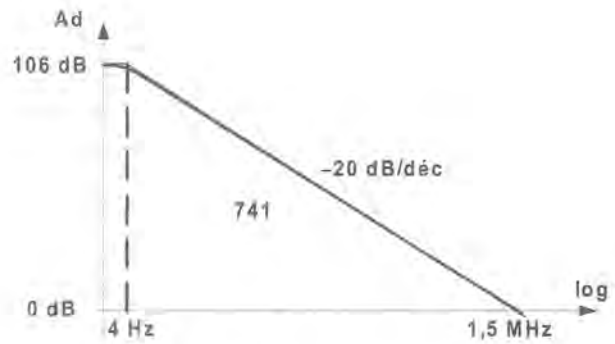


Figura 2.33. Ganancia diferencial en lazo abierto del amplificador operacional 741.

Si se considera el *ancho de banda* B como aquel conjunto de frecuencias en el que la variación de ganancia es menor de ± 3 dB[3], el ancho de banda del operacional en bucle abierto correspondería a $B_0 \approx 4$ Hz.

Sin embargo, el ancho de banda en lazo abierto no tiene demasiado interés ya que nuestros sistemas van a trabajar habitualmente en lazo cerrado y realimentados negativamente. Un sistema con ganancia A en lazo abierto y realimentado con un factor β tiene una ganancia en bucle cerrado, G determinada por [2.16]:

$$G = \frac{A}{1 + A\beta}$$

y, en el caso particular del operacional, resultaría ser:

$$G = \frac{A_d}{1 + A_d\beta} \tag{2.33}$$

[3] La definición del ancho de banda como el conjunto de frecuencias comprendido entre los valores cuya variación en ganancia sea de ±3 dB puede ser modificada y, en algunos casos, resulta poco conveniente. Una variación de 3 dB es equivalente a un cambio de ganancia de $\sqrt{2}$ que puede ser excesivo para algunas aplicaciones. En cualquier caso, se defina así o de otra forma, el lector debe saber lo que significa de forma cuantitativa ya que la aparente planitud de la curva de ganancia cuando se dibuja —como es habitual— en un diagrama logarítmico, puede sorprender al observar en un osciloscopio variaciones en las señales de hasta el 41% dentro del ancho de banda.

Para simplificar el primer caso de estudio, consideremos un operacional no inversor que sigue la expresión (2.33) y en el que la realimentación se ha efectuado sólo con resistencias ideales; siempre que $A_d >> 1/\beta$, se puede considerar que $1+A_d\beta \approx A_d\beta$ con lo que $G \approx 1/\beta$. En el diagrama de la figura 2.34 se han dibujado la ganancia diferencial en lazo abierto (igual que en la figura 2.33) y el inverso de la realimentación $1/\beta$. La anterior condición se cumple en la zona marcada como C, es decir, cuando la ganancia puede ser considerada como "infinita", mientras que cuando $A_d << 1/\beta$ (corresponde a la zona D de la figura 2.34), $1+A_d\beta \approx 1$ y $G \approx A_d$ y la ganancia final va cayendo igual que la del operacional en lazo abierto.

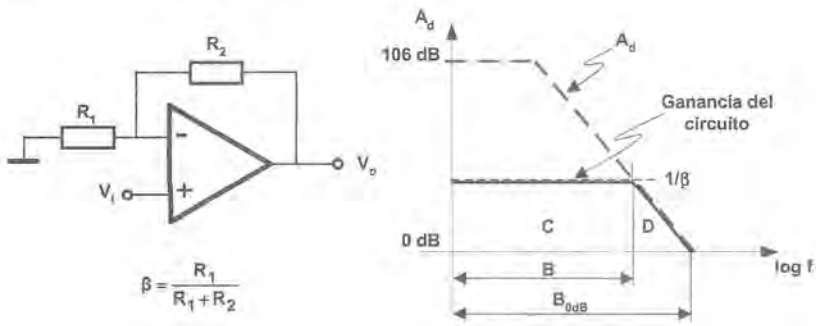


Figura 2.34. Circuito no inversor y respuesta en frecuencia resultante.

El ancho de banda del circuito amplificador no inversor resulta ser el indicado en la figura 2.34, mucho mayor que el correspondiente al amplificador sin realimentar (en lazo abierto). Lo mismo ocurre al realimentar cualquier sistema [2.9]. El valor de B se extiende hasta el punto en que $A_d = 1/\beta$, momento en el que la caída de ganancia es de 3dB, que corresponde a:

$$B = B_o(1 + A_d\beta) \tag{2.34}$$

Como puede observarse, el ancho de banda crece con el factor de realimentación con lo que, cuanto más realimentemos un circuito, mayor será su ancho de banda, es decir, a medida que bajamos la ganancia (la recta $1/\beta$ baja en la figura 2.34), más crece el ancho de banda. El límite está en que la recta $1/\beta$ no puede ser para ninguna frecuencia menor que 1 (0 dB) si la red de realimentación es resistiva, siendo el límite el caso de que $\beta = 1$, caso que corresponde al seguidor de emisor de la tabla 2.1. Para él se consigue del operacional el máximo ancho de banda B, que corresponde al punto en que $1/\beta = A_d = 0$ dB y que se denomina *ancho de banda a ganancia unitaria* (*unity gain bandwidth*) B_{0dB}. En función de ese valor, tendremos que el ancho de banda sería:

$$B = \beta B_o \tag{2.35}$$

EJERCICIO 2.11

Un circuito no inversor con ganancia 11 en continua (R_1 = 1K y R_2 = 10K) está construido con el operacional 741 que presenta la ganancia en lazo abierto que aparece en la figura 2.33. Determínese su ancho de banda. ¿Qué ocurriría si la ganancia fuese 110 en continua? ¿Y si fuese 1100?

SOLUCIÓN:

Si la ganancia es 11 en continua (a frecuencia 0 en la que se puede considerar que A_d es la máxima), significa que $1/\beta = 11$, con lo que el ancho de banda resultará ser unos 136 kHz. Para una ganancia de 110, bajaría hasta 13,6 kHz y para 1100, sólo sería de 1360 Hz.

La respuesta en frecuencia representada por la figura 2.33 es la propia de un tipo de amplificadores operacionales denominados *amplificadores compensados internamente en frecuencia* que se caracterizan porque la pendiente de la caída de la ganancia diferencial en lazo abierto es de –20 dB/déc hasta llegar, al menos, a 0dB.

El concepto de amplificador compensado en frecuencia hace referencia al comportamiento del dispositivo a la hora de ser realimentado que permite garantizar que no va a sobrepasar los márgenes de ganancia y fase derivados del Criterio de Nyquist [2.16] con realimentaciones puramente resistivas. Recordemos que el Criterio de Nyquist establece hasta qué punto se puede realimentar un sistema sin que llegue a aparecer el fenómeno de la realimentación positiva que originaría su oscilación; hay que tener en cuenta que el cambio progresivo de la fase de la tensión de salida respecto a la de entrada puede hacer que ambas se sumen en la entrada en vez de restarse (realimentación positiva) que sacaría al dispositivo de su funcionamiento estable.

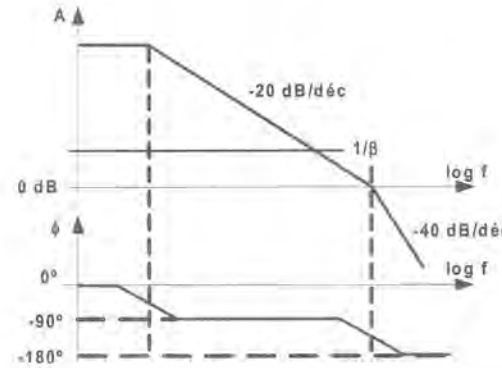


Figura 2.35. Diagrama de Bode asintótico de un operacional compensado internamente.

En la figura 2.35 se muestra un diagrama de Bode (ganancia y fase) de un amplificador compensado internamente en frecuencia realimentado con una red resistiva que fija un valor de $1/\beta$. Si nos fijamos en la expresión (2.13) para el caso en que $A_d\beta = -1$, tendremos que el denominador se hace nulo lo que significa que la ganancia es infinita o, lo que es lo mismo, para una entrada nula tendremos salida. Pues bien, si eso ocurre a una determinada frecuencia, tendremos un sistema que produce una salida de esa frecuencia sin necesitar entrada: estamos en presencia de un oscilador. Los osciladores son muy útiles en muchas ocasiones pero tienen una cierta tendencia a aparecer cuando menos se desea siguiendo la máxima de que "un circuito que no deba oscilar, entrará a la menor oportunidad en oscilación" [2.4].

Prevenir las oscilaciones no es tan difícil y un sistema bien calculado no entrará en oscilación ni el operacional aguardará "agazapado en su zócalo" para oscilar en cuanto miremos para otro sitio. Para un correcto diseño hay que evitar que el producto $A_d\beta$ llegue a valer –1 para alguna frecuencia, para lo que podemos dibujar ese producto en función de la frecuencia o, lo que es lo mismo observar el punto en que se cortan las curvas de $A_d(f)$ con $1/\beta(f)$: si su ganancia es mayor que 1 (0 dB) observaremos la fase y si no se provoca un –1, es decir, no llega a –180°, el denominador no se hará nulo y no oscilará; el *margen de fase* MF se define como la diferencia de fase hasta llegar a –180° de la curva $A_d\beta$ y determina lo lejos que está el sistema de la oscilación.

De la misma forma, el *margen de ganancia* MG establece cuánto se podría subir la ganancia en el punto en que la fase es –180° sin que se llegue alcanzar el valor $A_d\beta = -1$ que provocaría la oscilación.

En la figura 2.36a se muestra el valor de los márgenes de fase y ganancia para un operacional realimentado con una red resistiva pura.

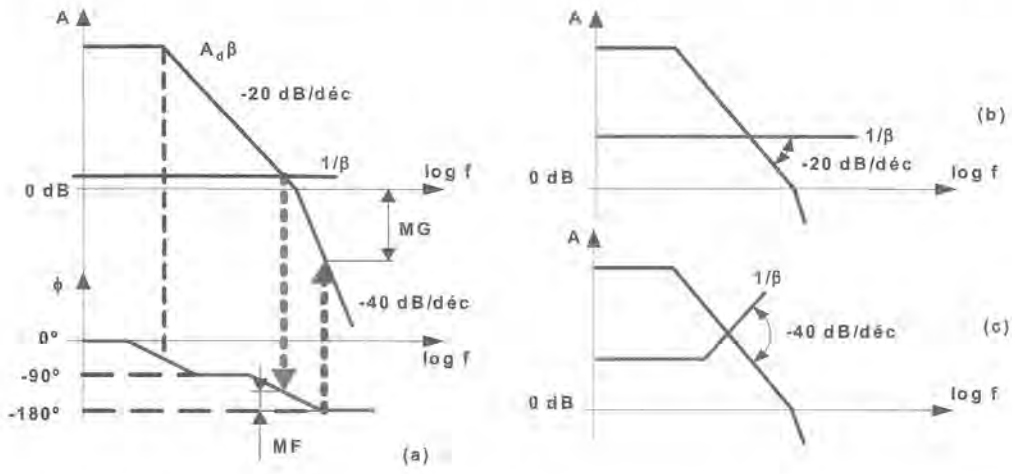


Figura 2.36. (a) Margen de ganancia y margen de fase; (b) sistema estable; (c) sistema inestable.

Trabajar en el límite suele ser peligroso porque las tolerancias de los componentes y las derivas de cualquier tipo pueden poner al sistema más allá del punto crítico y tendríamos un oscilador haciendo cierta la máxima que antes decíamos, así que se trabaja con un margen de fase mínimo que nos garantiza una cierta seguridad de no-oscilación; este margen de fase suele ser de 45° lo que implica que las curvas de A_d y $1/\beta$ se corten con pendiente relativa menor o igual a 20 dB/déc. En la figura 2.36b se muestra un sistema estable ya que la pendiente relativa es de 20 dB/déc mientras que en la figura 2.36c se muestra un sistema en el que la pendiente relativa entre ambas curvas es de 40 dB/déc lo que significa ¡peligro de oscilación!

Observando la figura 2.36b se puede ver que los amplificadores compensados internamente en frecuencia no oscilarán nunca con realimentación por redes resistivas puras ya que el caso más desfavorable es el del seguidor de emisor ($\beta = 1/\beta = 1$) que cortaría en el límite del margen de fase establecido por seguridad.

Los amplificadores operacionales compensados internamente en frecuencia se construyen con una compensación de fase en la etapa intermedia que reduce su ancho de banda pero garantiza la estabilidad incluso con realimentación unitaria. Para todos ellos, considerando el caso de un amplificador no inversor, la ganancia y el ancho de banda se contraponen según pudimos comprobar en el ejemplo 2.11 y su producto se mantiene constante en toda la zona estable (la zona en la que la pendiente cae a –20 dB/déc en A_d) que define el producto *ganancia-ancho de banda (gain-bandwidth)* o GBW como una característica del operacional expresada en Hz y que combina su capacidad de amplificación y su ancho de banda.

Los operacionales compensados internamente en frecuencia suelen tener anchos de banda reducidos y si se desea una gran amplificación a una frecuencia elevada, el valor de GBW actúa como una limitación insalvable. Para estos casos, se utilizan operacionales no compensados en frecuencia que no incluyen ninguna compensación en sus etapas intermedias y que presentan una característica de ganancia en lazo abierto como la que se muestra en la figura 2.37a.

La zona útil de este tipo de amplificadores corresponde a la zona en la que la pendiente de A_d cae a –20 dB/déc en donde una realimentación resistiva pura permitiría una ganancia muy elevada y un amplio ancho de banda (figura 2.37b), mucho mayor que el que tendría si se usase uno compensado (figura 2.37c):

el peligro de estos operacionales está en su uso para ganancias bajas (1/β bajas) ya que nos podemos encontrar con inestabilidad según se muestra en la figura 2.37d en que el corte de las curvas de A_d y 1/β se produce con pendiente relativa de 40 dB/déc.

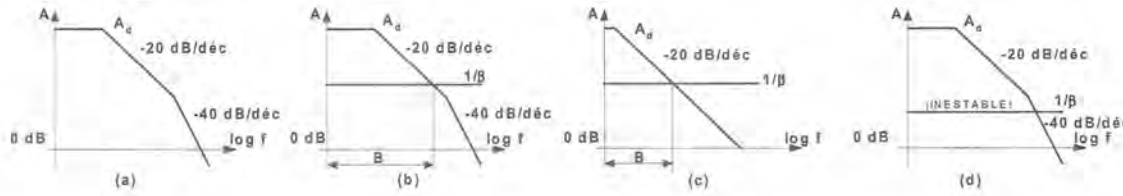


Figura 2.37. (a) Ganancia en bucle abierto de un operacional no compensado; (b) ancho de banda de un operacional no compensado realimentado mucho mayor que el que tendría si no fuese compensado (c); (d) sistema inestable.

Aun con amplificadores no compensados, se puede trabajar con ganancias bajas y añadir la fase necesaria al sistema para conseguir que entre dentro de los límites de estabilidad (figura 2.38a) o compensar el propio operacional añadiendo un polo adicional que modifique su respuesta frecuencial (figura 2.38b). Lo primero se puede hacer con una red exterior de adelanto o de adelanto-retraso de fase mientras que lo segundo se suele hacer mediante un condensador exterior al operacional situado entre dos puntos de la etapa intermedia accesibles en su patillaje (en cada caso, se recurrirá a las especificaciones del fabricante).

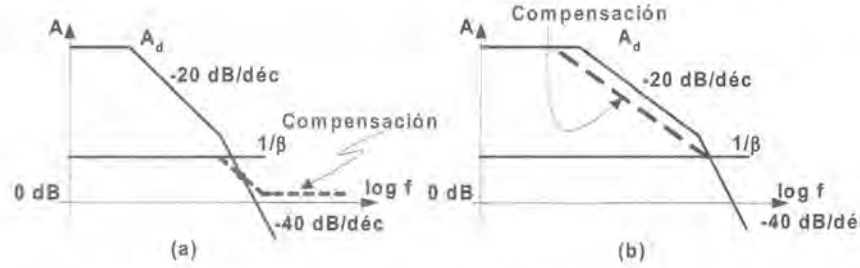


Figura 2.38. Compensación de operacionales: (a) añadiendo fase; (b) modificando la respuesta en lazo abierto del dispositivo.

Como norma general, es mejor utilizar amplificadores compensados cuando se pretenden obtener ganancias bajas y acudir a amplificadores no compensados cuando es preciso un valor del producto ganancia-ancho de banda muy alto.

Pero el operacional no sólo responde a la tensión diferencial: hemos visto que también es capaz de amplificar la tensión de modo común aunque con bajos valores de ganancia que suelen quedar definidos por la CMRR. Nos podemos preguntar cómo responde el operacional en frecuencia desde el punto de vista de la CMRR ya que nos interesará que su valor se mantenga lo más alto que sea posible en toda la banda de trabajo.

Pues bien, la CMRR es alta —en cada operacional— para frecuencias bajas mientras que va cayendo a medida que aumentamos la frecuencia para volver a subir cuando las frecuencias son muy elevadas. Este comportamiento hace que el proceso de reducción de interferencias de modo común sea muy efectivo para bajas frecuencias (interferencias de red a 50 Hz como el caso del ejemplo 2.10) pero no lo sea tanto cuando

la interferencia de modo común es de una frecuencia alta. En la figura 2.39 se muestra un gráfico en el que aparece el comportamiento de la CMRR en función de la frecuencia con las limitaciones de que hablamos.

Figura 2.39. Variación de la CMRR con la frecuencia.

Del mismo modo, la razón de rechazo de la alimentación PSRR tampoco tiene buenas características en alta frecuencia con lo que las variaciones en la tensión de alimentación provocadas por ruidos tendrán bastante facilidad para aparecer en la salida del operacional formando parte de v_o. En la figura 2.40a aparece el comportamiento de la PSRR en función de la frecuencia observando que su valor va cayendo a medida que sube la frecuencia haciendo que los circuitos con operacionales sean bastante sensibles al ruido en las alimentaciones [2.12]. En la figura 2.40b se muestra cómo reducir el efecto de las perturbaciones de alta frecuencia a costa de filtrar —en las proximidades del operacional— las alimentaciones.

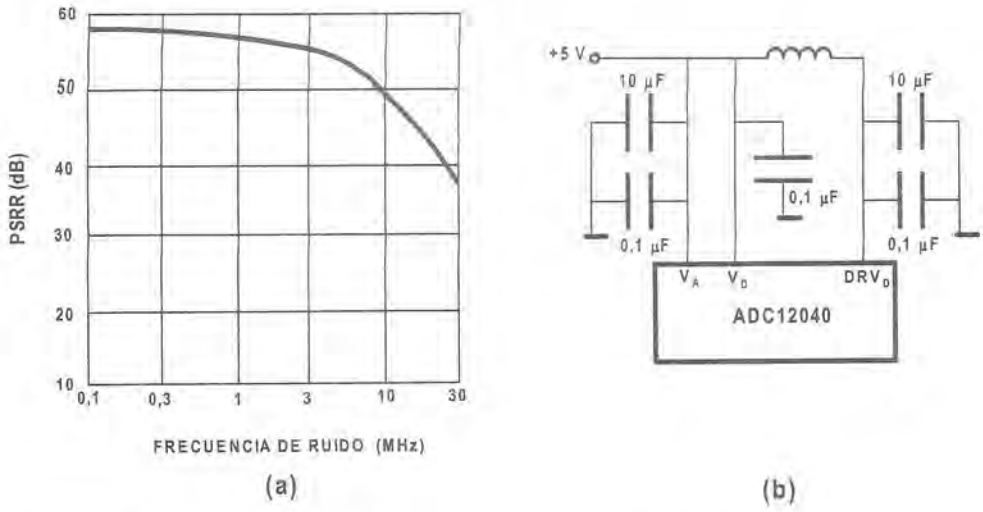


Figura 2.40. (a) Variación de la PSRR con la frecuencia y, (b) filtro en la alimentación para eliminar los ruidos de alta frecuencia de la alimentación.

EJERCICIO 2.12

Se pretende amplificar la señal procedente de un sensor que produce en su salida una tensión que puede variar entre 0,00 y 10,0 mV de pico y cuya frecuencia máxima puede ser 10 kHz. Si se pretende obtener una señal de salida cuyo valor máximo sea 2,50 V de pico, ¿cómo se especificaría el operacional desde el punto

de vista de su comportamiento frecuencial para hacerlo en una sola etapa de amplificación? ¿Y si la máxima frecuencia fuera de 1 MHz?

SOLUCIÓN:

Si utilizamos un amplificador no inversor para realizar esta función, su ganancia debería ser de 2,50/0,0100 = 250 con lo que $(R_1+R_2)/R_1 = 250$ (48,0 dB) y si elegimos $R_1 = 1K$, R_2 sería de 249K. En estas condiciones, el valor del producto GBW = 250 × 10 kHz = 2,5 MHz que no resulta un valor excesivo para un operacional compensado en frecuencia y será fácil encontrar uno que nos sirva.

En el segundo caso, GBW = 250 × 1 MHz = 250 MHz que resulta un valor muy alto. Habría que recurrir a un amplificador no compensado con casi total seguridad. En estas condiciones, habrá que cuidar que no aparezcan situaciones de inestabilidad.

A la hora de elegir el operacional para este segundo caso, nos fijaremos en la frecuencia del segundo polo que define el momento a partir del cual la pendiente de caída de A_d es de 40 dB/déc. Ésa será la máxima frecuencia a la que podrá trabajar el amplificador realimentado con resistencias y sin ningún tipo de compensación (figura 2.41).

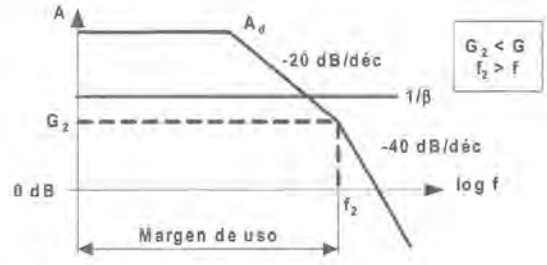


Figura 2.41. Condiciones para emplear el amplificador del ejercicio 2.12.

Las condiciones que se deben cumplir son:

a) La frecuencia del segundo polo debe ser mayor que la máxima frecuencia que precisemos.

b) La ganancia en lazo abierto en ese punto —el segundo polo— debe ser menor que la necesaria para el caso que nos ocupe.

Cuando se llega a este punto se suele decir que uno de los objetivos que se persiguen en el desarrollo de nuevos amplificadores operacionales es lograr valores más altos de GBW y es normal ver anuncios de fabricantes que —a bombo y platillo— presentan nuevos componentes con excepcionales valores del producto ganancia-ancho de banda. Parece pues, que cuanto más grande sea el ancho de banda, mejor. Esta aseveración es cierta porque la capacidad de amplificar señales de alta frecuencia es algo que se desea en muchas ocasiones y que no es fácil de lograr. Pero eso no significa que **siempre** haya que disponer de grandes anchos de banda; es más, veremos que, casi siempre, un ancho de banda excesivo es contraproducente.

Slew-rate

El ancho de banda mide hasta qué frecuencia se puede llevar a trabajar a un circuito dado, pero en condiciones de *pequeña señal*, es decir, para señales de muy bajo valor en comparación con la máxima tensión que se podría desarrollar en la salida. Sin embargo, cuando los valores de la tensión de salida son grandes, es posi-

ble que nos encontremos con la limitación impuesta por la máxima velocidad a la que la salida puede variar de valor y que viene definida por el *slew-rate*, SR que suele medirse en V/μs y que determina la máxima pendiente que la tensión de salida puede presentar.

El valor del *slew-rate* viene determinado por los mismos factores que limitan el ancho de banda y, por lo tanto, tendrá relación con éste: en amplificadores compensados en frecuencia, su origen está en el condensador que se sitúa en la etapa intermedia (ver figura 2.11c) para efectuar esa función, mientras que en los no compensados se debe a las diversas capacidades parásitas que formarán parte inevitable de la topología. En el primer caso tenemos amplificadores operacionales de poco ancho de banda y de *slew-rate* bajo mientras que en el segundo caso los operacionales tendrán un elevado ancho de banda y un alto valor de *slew-rate*.

Si —en cualquier topología con operacionales— introducimos una señal de tal forma que tengamos una tensión senoidal a la salida de valor máximo V, tendremos que la señal se reproducirá correctamente si no se produce en ningún momento una variación de tensión más rápida que el SR; así, si $v_o = V\text{sen}(2\pi ft)$, tenemos que la variación de v_o vendrá dada por su derivada:

$$\frac{dv_o}{dt} = 2\pi f V\cos(2\pi f t) \Rightarrow \left(\frac{dv_o}{dt}\right)_{máx} = 2\pi f V$$

Por tanto, para que la tensión de salida aparezca correctamente, se tendrá que cumplir que:

$$2\pi f V < SR \Rightarrow f_{máx} = \frac{SR}{2\pi V} \tag{2.36}$$

lo que impone una frecuencia máxima y viene a representar otra limitación para el ancho de banda del circuito que resulta independiente de la realimentación, que depende única y exclusivamente del operacional utilizado y que afecta por igual a todos los circuitos que se construyan usando como circuito básico ese operacional. Como se desprende de la expresión (2.36), la frecuencia máxima también depende de la tensión de salida, disminuyendo a medida que la tensión de salida es más alta. Esta limitación de frecuencia impone un ancho de banda adicional al operacional que se puede calificar como *ancho de banda de potencia* (*power bandwidth*) que termina de definir la zona de uso de cualquier operacional en un plano v_o-f según se indica en la figura 2.42.

Figura 2.42. Límites de uso de un operacional.

El efecto del *slew-rate* se puede cuantificar en términos de distorsión armónica ya que si se supera el límite que impone, modifica la distribución de armónicos de la tensión de salida aunque el efecto es diferente según el tipo de onda de que se trate:

a) Una onda senoidal sólo sufre distorsión si se supera el *slew-rate* con lo que para frecuencias menores que la impuesta por la expresión (2.36), la distorsión armónica por este concepto será nula. Sin embargo, si no se cumple (2.36), la tensión de salida variará a la velocidad impuesta por el SR mientras la derivada de v_o supere el valor de SR.

En la figura 2.43 se muestran ambos casos para el seguidor de emisor construido con un operacional compensado internamente en frecuencia, indicando la distribución de armónicos de la señal a la salida. Si se supera el *slew-rate*, el armónico fundamental de la tensión de salida baja de valor y aparecen más armónicos en la tensión de salida incrementándose el valor de THD.

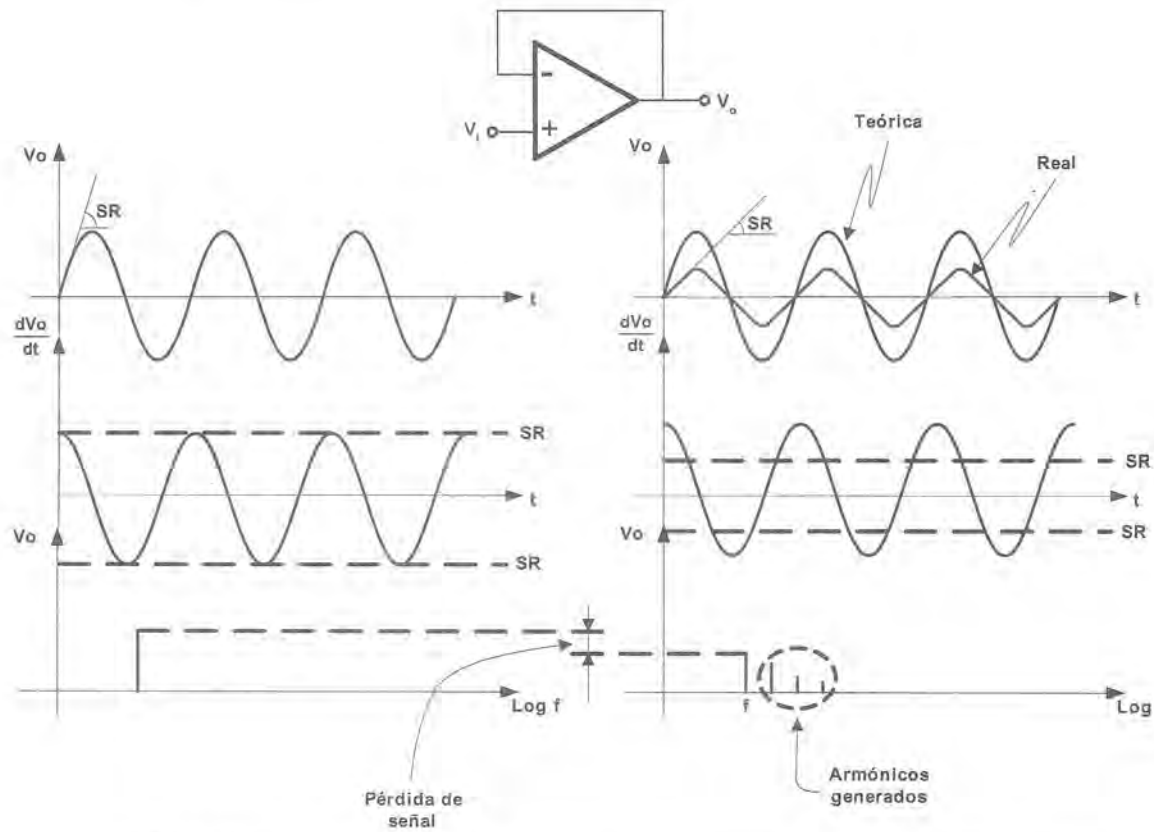


Figura 2.43. Efecto del *slew-rate* sobre la salida de un operacional cuando se trabaja dentro del ancho de banda de señal: a la izquierda, cuando la señal es suficientemente pequeña, la reproducción es correcta, pero con valores excesivos, no es posible alcanzar el valor teórico y el operacional "patina" (figura de la derecha), produciéndose pérdida de señal y generación de distorsión armónica.

b) Una onda cuadrada siempre sufre el efecto del *slew-rate* ya que para reproducir el flanco teórico de subida se necesitaría un valor infinito de SR. El efecto puede cuantificarse mediante un recorte en la distribución de armónicos de alta frecuencia según se indica en la figura 2.44 y que no corresponde al recorte impuesto por el ancho de banda de pequeña señal ya que depende también de la amplitud de v_o.

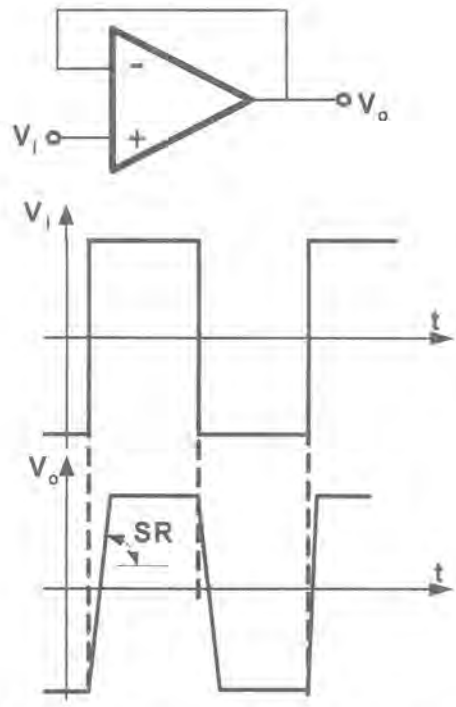


Figura 2.44. Efecto del *slew-rate* sobre una onda cuadrada.

Los valores de *slew-rate* dependen mucho del tipo de operacional de que se trate pero, en general, son bajos para los compensados (para el 741, SR = 0,5 V/μs y para el TL081, SR = 13 V/μs) y altos para los no compensados internamente en frecuencia.

EJERCICIO 2.13

Supongamos que queremos construir un amplificador no inversor de ganancia 10 usando el 741 cuyos parámetros característicos en frecuencia son GBW = 1 MHz y SR = 0,5 V/μs y que está alimentado simétricamente con $+V_{cc}$ = 12 V y V_{ee} = –12 V. ¿Cuál es la máxima frecuencia de una señal de 1,00 V de amplitud de entrada que se podría amplificar sin distorsión? ¿Cuál es la máxima amplitud de una señal de 10 kHz que se podría amplificar sin distorsión? ¿Cómo sería la tensión de salida cuando introducimos una señal cuadrada de 10 kHz y 1,00 V de amplitud en la entrada?

SOLUCIÓN:

Con la alimentación de que se dispone, el operacional podría dar tensiones de ±10 V en la salida sin dificultad, así que, si sólo nos fijamos en el producto ganancia-ancho de banda y teniendo en cuenta la ganancia del circuito, el 741 sería capaz de amplificar señales de hasta 100 kHz con ganancia 10 pero si diésemos esa respuesta sería incorrecta por la limitación impuesta por el *slew-rate*. Según la ecuación (2.36) y con 10,0 V en la salida, tenemos que la máxima frecuencia de salida sería:

$$f_{máx} = \frac{0{,}5 \text{ V/}\mu\text{s}}{2\pi 10} = 7{,}96 \text{ kHz}$$

con lo que cualquier señal de más frecuencia sufriría distorsión ya que el operacional no podría seguirla en los momentos en que el cambio de tensión es más rápido que el *slew-rate*.

Si tratamos de responder a la segunda pregunta, nos encontramos que, despejando la tensión de la ecuación (2.36) y sustituyendo los valores, tenemos que:

$$V_{máx} = \frac{0{,}5\ V/\mu s}{2\pi 10 \cdot 10^3\ Hz} = 7{,}95\ V$$

Este será el mayor valor de tensión que el circuito podrá tener en la salida; por tanto, la máxima amplitud de la señal de entrada podrá ser de 0,795 V.

Finalmente, para conocer la tensión de salida en el tercer caso, tengamos en cuenta que el circuito multiplica idealmente por 10 la entrada, es decir, en la salida deberíamos tener una señal cuadrada de 10V de amplitud. Sin embargo, en los cambios de nivel, la salida no podrá variar más rápidamente que el *slew-rate* y, como tiene que pasar de –10 a +10 V y viceversa y SR = 0,5 V/μs, invertirá un tiempo de 20/0,5 = 40 μs para subir y otro tanto para bajar, lo que representa el 80% del periodo total de la onda que tendrá un aspecto como el de la figura 2.45.

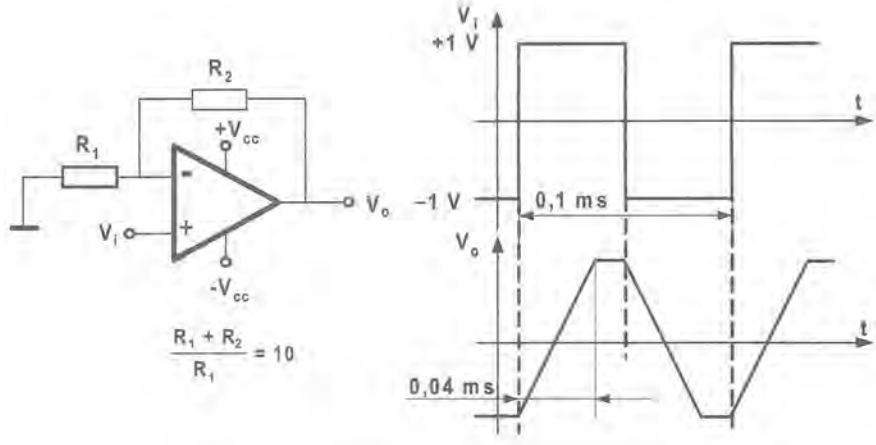


Figura 2.45. Circuito y formas de onda del ejercicio 2.13.

El efecto del *slew-rate* sobre las tensiones de salida de los operacionales no puede compensarse y la única solución sería la de acudir a un operacional de mejores características en este parámetro. En cualquier caso, es importante conocer las limitaciones que impone en el funcionamiento de circuitos con operacionales para saber a qué debemos atenernos en nuestro diseño.

2.2.3.3 Ruido interno

Como ya hemos analizado en el primer apartado de este capítulo, uno de los principales problemas que aparecen en todos los sistemas de instrumentación es la presencia de ruido, es decir, señales no deseadas que se superponen a la señal que nos interesa y que tienden a ocultar parcialmente la información que contiene. En general, la importancia del ruido se mide con la relación señal/ruido ya definida y, como quiera que los sistemas de instrumentación se caracterizan por tener señales débiles —al menos en sus entradas— la importancia relativa del ruido será mayor que la que pueda tener en otros sistemas electrónicos.

A los circuitos de amplificación les llega una señal que —como su nombre indica— deberán amplificar; esa señal podrá tener una componente de ruido que será también amplificada junto con la señal pero, bajo condiciones ideales de amplificación lineal sin distorsión de ninguna clase, la relación S/N se mantendrá a la salida.

Desafortunadamente, eso no es cierto.

El amplificador no resulta ideal tampoco en este aspecto y añadirá una porción adicional de ruido a la señal con lo que, en las condiciones anteriores, tendremos una relación señal/ruido peor a la salida que a la entrada. Ese ruido puede tener dos orígenes:

- Ruido interno, generado por el propio circuito.
- Interferencias externas que se añaden desde circuitos o sistemas externos.

En este capítulo trataremos sólo del primero de estos orígenes, dejando para el capítulo de interferencias electromagnéticas el segundo.

Cuando observamos la tensión de salida de un operacional podemos comprobar que aun con todas las entradas conectadas a masa, con todos los ajustes de tensiones de desviación correctamente efectuados para corregir sus efectos, se desarrolla un nivel de tensión que contiene un gran número de frecuencias y que se conoce como *nivel de ruido basal* y corresponde al menor nivel de señal que puede discriminarse en la salida sin ningún tipo de procesamiento estadístico.

Las señales de ruido suelen contener un gran número de frecuencias por lo que no cabe hablar de señales de un valor determinado, sino que siempre se expresa en términos de densidad espectral, utilizando unidades de $V/\sqrt{Hz}$ o de $A/\sqrt{Hz}$ y que, para conocer valores de tensión o corriente no queda más remedio que definir una banda de frecuencias determinada.

El ruido que se genera en un circuito electrónico puede tener diversos orígenes. En general, se puede hablar de:

a) *Ruido térmico*, también conocido como ruido Johnson o ruido blanco[4] por las características espectrales que presenta: una distribución espectral plana lo que significa que **se produce a todas las frecuencias**. Este ruido se produce en cualquier elemento de comportamiento resistivo, se debe a la agitación térmica de las partículas y crece con la temperatura y con el valor de la resistencia equivalente del componente; matemáticamente, se modela mediante una fuente de tensión de valor eficaz V_t en serie con la propia resistencia R donde se genera [2.18]:

$$V_t^2 = \int_{f_i}^{f_s} 4kTRdf \qquad (2.37)$$

donde $k = 1{,}38 \cdot 10^{-23}$ J/K es la constante de Bolztman, T es la temperatura en grados Kelvin y f_i y f_s son las frecuencias de corte inferior y superior del circuito, respectivamente. Como se produce a todas las frecuencias y, en consecuencia, crece con el ancho de banda, los circuitos que se diseñen deberán restringir su ancho de banda al necesario para amplificar la señal ya que cualquier ancho de banda mayor no mejorará la señal e introducirá más ruido, empeorando la relación S/N.

El ruido térmico presenta una distribución normal según puede observarse en la figura 2.46.

[4] El nombre de "ruido blanco" hace referencia a una similitud con la luz: una luz que contenga todas las longitudes de onda (todas las frecuencias) es blanca. Aquí, el ruido contiene todas las frecuencias y, de ahí, su calificativo. El nombre de "ruido rosa" hace referencia a un ruido que sólo contiene frecuencias bajas, también por similitud con la luz: una luz que contenga sólo frecuencias bajas (longitudes de ondas altas) no contiene colores azulados o verdosos y tiene un tono rojizo. Este tipo de ruido baja con la frecuencia a razón de 1/f. Existen más asociaciones de colores con distribuciones espectrales de ruidos, aunque tienen menos importancia en amplificadores operacionales.

Figura 2.46. Distribución estadística del ruido térmico.

b) *Ruido de parpadeo*, ruido *flicker* o ruido 1/f que es un ruido con ancho de banda limitado (ruido rosa) y que aparece en todos los componentes activos y en las resistencias de carbón. Su origen es diverso pero siempre se asocia a una corriente continua circulante por el componente y se modela mediante una fuente de tensión o una fuente de corriente:

$$V_f^2 = \int_{f_i}^{f_s} \frac{k_e^2}{f} df \qquad i_f^2 = \int_{f_i}^{f_s} \frac{k_i^2}{f} df \tag{2.38}$$

donde las frecuencias f_i y f_s son, como antes, los límites del ancho de banda y las k son proporcionales a I, siendo I la corriente continua [2.16]. La presencia de la frecuencia en el denominador hace que la importancia de este ruido sea menor en frecuencias elevadas y mayor en las bajas.

c) *Ruido de golpeteo (shot noise)*, que se ocasiona por el movimiento aleatorio de los electrones al atravesar cualquier barrera de potencial y que presenta una distribución espectral plana (es un ruido blanco), creciendo con la corriente que circule por el componente; se modela como una fuente de corriente de valor I_s de valor:

$$I_s^2 = \int_{f_i}^{f_s} 2qI_m df \tag{2.39}$$

donde $q = 1{,}6 \cdot 10^{-19}$C es la carga del electrón e I_m es la corriente media. Como se puede observar, el ruido crece con la corriente media y, como siempre, con el ancho de banda del sistema.

d) *Burst noise*, que se origina por impurezas en las uniones p-n a bajas frecuencias (menores de 100Hz).

e) *Ruido de transición*, que se produce cuando los portadores de carga tienen varios caminos para avanzar como en el caso de los transistores bipolares.

f) *Ruido de avalancha*, que se ocasiona en las uniones que trabajan cerca de la zona de avalancha por los choques de los electrones con la red que arrancan nuevos pares electrón-hueco de forma aleatoria provocando pulsos de corriente.

En los amplificadores operacionales, el ruido generado internamente se modela mediante una fuente de tensión en la entrada y dos fuentes de corriente —una en cada una de las entradas— que incluyen la participación del ruido de origen térmico, del ruido de golpeteo y del ruido de *flicker*. En la figura 2.47 se pueden observar tres modelos para tener en cuenta el ruido en un operacional cualquiera [2.16], [2.7].

Las fuentes de ruido de los tres modelos están relacionadas entre sí y podrá usarse cualquiera de ellos en función de los datos que proporcione el fabricante del operacional que estemos usando. Se supone que los datos que facilita son la mejor forma de modelar el comportamiento del operacional en este aspecto.

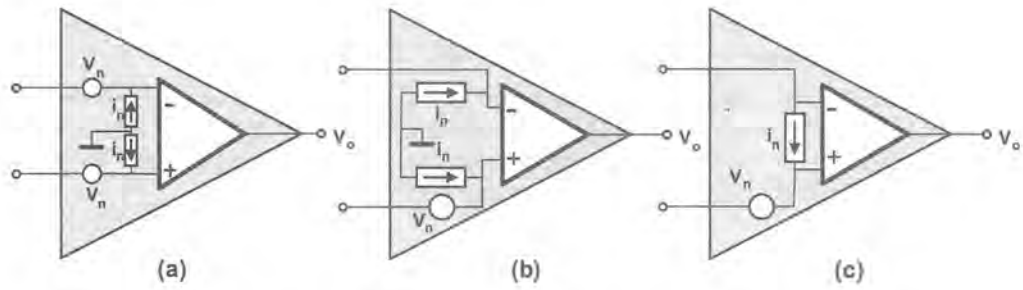


Figura 2.47. Modelos de ruido interno en un operacional.

Por otro lado, sea cual sea el origen del ruido interno, se podrá considerar que es aditivo puesto que:

$$V_t^2 = \int_{f_i}^{f_s} (v_1 + v_2)^2 df = \int_{f_i}^{f_s} v_1^2 df + \int_{f_i}^{f_s} v_2^2 df + \int_{f_i}^{f_s} v_1 v_2 df = V_1^2 + V_2^2$$

ya que el tercer sumando de la integral es nulo al representar el producto de dos variables aleatorias de media nula y donde v_i representa el valor instantáneo del ruido para cada frecuencia y V_i denota su valor eficaz. La anterior expresión puede extenderse a más señales de ruido, obteniendo una expresión más genérica:

$$V_t^2 = \sum_{i=1}^{n} V_i^2 \tag{2.40}$$

Considerando la suma de todos los ruidos de origen interno en el operacional, las fuentes de los modelos de la figura 2.47 representarán todos estos efectos.

La importancia relativa de las fuentes de ruido del modelo del operacional es un factor a tener en cuenta ya que el efecto de las tensiones de ruido suele ser mayor que el de las corrientes de ruido y suele usarse como modelo de ruido el modelo que corresponde sólo a la fuente de tensión de ruido en la entrada no inversora del operacional. En la figura 2.48 se muestra este modelo reducido que sirve para un gran número de aplicaciones y que es muy habitual en los circuitos de audio.

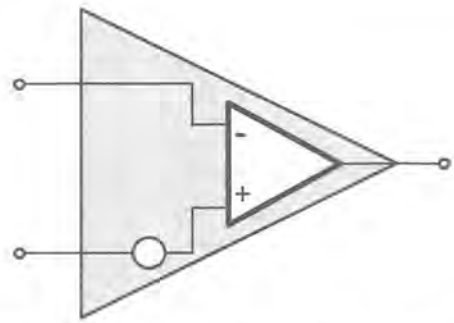


Figura 2.48. Modelo elemental de ruido en un operacional.

Si lo que tenemos es un convertidor corriente/tensión como el que se usa en las etapas de tratamiento inicial de gran cantidad de sensores cuya salida es una corriente, lo dicho antes debe ser matizado ya que podemos encontrarnos con valores muy altos de la resistencia de realimentación, mientras que el propio sensor se comportará como una fuente de corriente (alta impedancia) ocasionando valores muy altos de impedancia con lo que el efecto de la corriente de ruido será importante y como quiera que la ganancia de

tensión de un convertidor I/V es unitaria, la tensión de ruido aparecerá sin amplificar en la salida y su efecto será mucho menor que en el caso de amplificadores de tensión.

En cualquier caso, los fabricantes suelen proporcionar los datos de las tensiones y las corrientes de ruido que deben considerarse y que tienen un aspecto como el que se aprecia en la figura 2.49 aunque no debemos sorprendernos si no disponemos de la información correspondiente a la corriente de ruido ya que hay muchas hojas de características que no la incluyen por el motivo antes indicado.

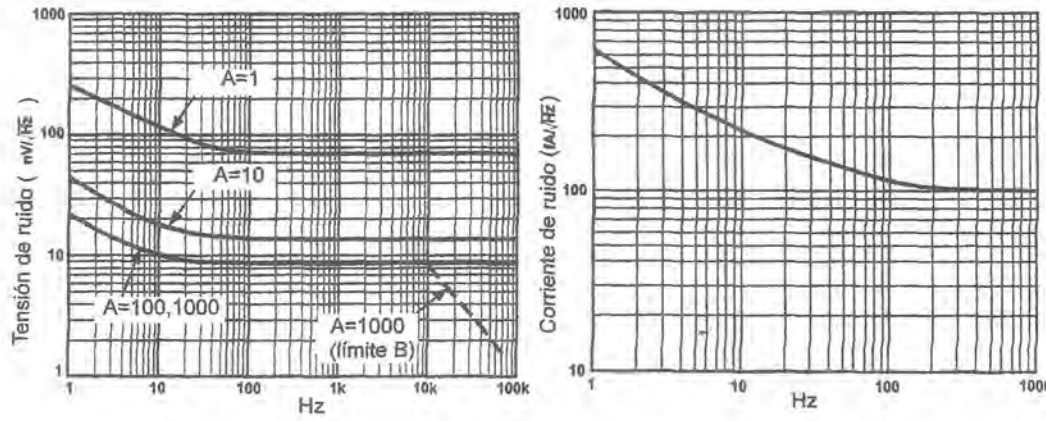


Figura 2.49. Tensión y corriente de ruido del amplificador AD620 de *Analog Devices* (tomado de las hojas de características del fabricante).

En la figura 2.50 se observa que el aspecto de la corriente y la tensión de ruido son similares, teniendo una zona de baja frecuencia en donde predomina el *flicker* (ruido 1/f) y una zona de alta frecuencia en donde predomina el ruido blanco cuyo origen es el térmico y el golpeteo.

Las gráficas de la figura 2.50 son poco claras, resultan difíciles de interpretar y la información que proporcionan debe ser presentada de una forma más cómoda y más comprensible para el usuario. Para ello, consideremos sólo la tensión de ruido y tengamos en cuenta que los ruidos de cada uno de los orígenes son aditivos; entonces, si V_f es el valor eficaz de la tensión de ruido *flicker* y V_w es la tensión eficaz de ruido blanco (debido al golpeteo y al ruido térmico), tenemos que la tensión eficaz de ruido será:

$$V_n^2 = V_f^2 + V_w^2$$

El ruido blanco tiene densidad constante C, mientras que el *flicker* sigue la expresión 1/f, al hacer las integrales respectivas queda:

$$V_w^2 = \int_{f_i}^{f_s} C\,df \qquad y \qquad V_f^2 = \int_{f_i}^{f_s} \frac{K^2}{f}\,df \Rightarrow V_n^2 = C\left(f_s - f_i\right) + K^2 \ln\frac{f_s}{f_i}$$

con $C = K^2/f_{ncv}$ y siendo f_{ncv} la frecuencia a la que se igualan las densidades espectrales del ruido blanco y del *flicker* (figura 2.50).

En esas condiciones, la tensión eficaz de ruido será:

$$V_n = \sqrt{\left[C\left(f_s - f_i\right) + f_{ncv} \ln\frac{f_s}{f_i}\right]} \qquad (2.41)$$

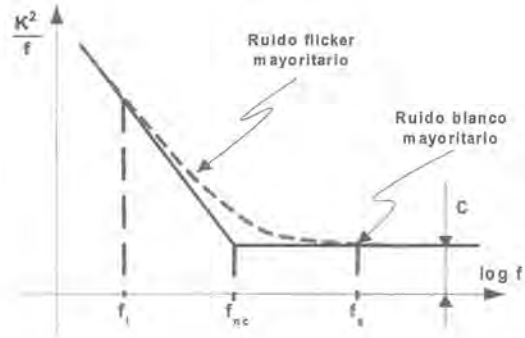


Figura 2.50. Las dos causas más importantes de ruido en un operacional y el ruido total generado.

expresión que resulta muy cómoda de utilizar ya que todos los parámetros tienen una lectura muy fácil en la gráfica de la tensión de ruido (figura 2.50).

Sin embargo, para realizar un análisis más exacto del ruido de un sistema amplificador, deberíamos tener en cuenta todos los parámetros de ruido del sistema, lo que incluiría también las resistencias que formen parte del circuito. Si consideramos el circuito de un amplificador diferencial (figura 2.51a) y pretendemos calcular el ruido producido, usaremos el modelo de la figura 2.51b en donde R_3 es la resistencia que se ve desde la entrada no inversora, es decir, el paralelo de R_1 y R_2.

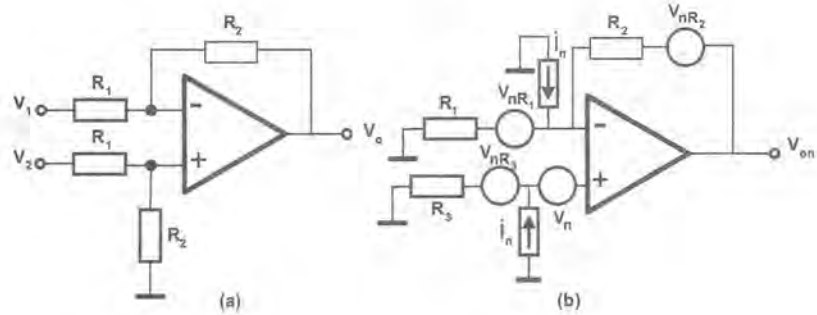


Figura 2.51. El ruido en un amplificador diferencial: (a) circuito; (b) circuito teniendo en cuenta todas las fuentes de ruido (el propio operacional y las resistencias).

El análisis se realizará por superposición, suponiendo un comportamiento lineal del circuito y considerando aditivos los ruidos según (2.40). La tensión de ruido debido a las resistencias V_{oR} será, entonces:

$$V_{oR} = \sqrt{\frac{R_2}{R_1}\int_{f_i}^{f_s} 4kTR_1 df + \frac{R_1+R_2}{R_1}\int_{f_i}^{f_s} 4kTR_3 df + \int_{f_i}^{f_s} 4kTR_2 df} \qquad (2.42)$$

es decir, operando tenemos que:

$$V_{oR} = 2(R_2+R_1)\sqrt{\frac{kT(f_s-f_i)}{R_1}} \qquad (2.43)$$

Por otro lado, teniendo en cuenta el ruido debido al operacional, tenemos que es la suma de los ruidos debidos a las distintas fuentes de ruido. El ruido debido a la fuente de tensión será el calculado en (2.41) por lo que, en la salida, aparecerá:

$$V_{oV} = \frac{R_1 + R_2}{R_1}\sqrt{C(f_s - f_i) + f_{ncu}\ln\frac{f_s}{f_i}} \tag{2.44}$$

siendo f_{ncu} la frecuencia a la que se igualan el ruido térmico y el *flicker* en la curva de tensión de ruido. El ruido debido a las fuentes de corriente será, para cada una de ellas:

$$\begin{aligned} V_{oI1} &= R_3\frac{R_1 + R_2}{R_1}i_w\sqrt{(f_s - f_i) + f_{nci}\ln\frac{f_s}{f_i}} \\ V_{oI2} &= R_2 i_w\sqrt{(f_s - f_i) + f_{nci}\ln\frac{f_s}{f_i}} \end{aligned} \tag{2.45}$$

Así, para el caso del amplificador diferencial, en que R_3 es el paralelo de R_1 y R_2, la suma de los efectos de las corrientes resultará ser:

$$V_{oI} = \sqrt{V_{oI1}^2 + V_{oI2}^2} = \sqrt{2}R_2 i_w\sqrt{(f_s - f_i) + f_{nci}\ln\frac{f_s}{f_i}} \tag{2.46}$$

Entonces, el resultado del ruido total V_{on} que se produce a la salida del amplificador es:

$$V_{on} = \sqrt{V_{oR}^2 + V_{oV}^2 + V_{oI}^2}$$

Sustituyendo las expresiones (2.43), (2.44) y (2.45), se obtiene:

$$V_{on} = \frac{R_1 + R_2}{R_1}\sqrt{\left[4kTR_1 + C + 4i_w^2R_3^2\right](f_s - f_i) + \left[Cf_{ncv} + 4R_3^2i_w^2f_{nci}\right]\ln\frac{f_s}{f_i}} \tag{2.47}$$

En general, de la observación de las expresiones (2.43), (2.44), (2.46) y (2.47) se puede concluir que la importancia del ruido crece siempre con el valor de R_2 y esto se hace especialmente importante en el caso de las corrientes de ruido; el efecto de las resistencias R_1 es menos importante, también en líneas generales, luego habrá que cuidar el ruido en aquellos circuitos que tengan elevados valores de R_2, es decir:

- Circuitos con ganancia elevada.
- Circuitos de conversión I/V en que sólo tienen R_2, que es la que determina la ganancia.

También se puede concluir que la inclusión de resistencias en la entrada no inversora para tareas de compensación de los efectos de la corriente de polarización va a llevar consigo el correspondiente incremento del efecto de la corriente de ruido según indica la primera de las expresiones (2.45).

Una forma de cuantificar el conjunto de los efectos frecuenciales no deseados introducidos por el operacional puede ser el parámetro de distorsión armónica más ruido (THD+N) que incluye tanto los armónicos introducidos por el paso de amplificación como las tensiones de ruido y que se suele dar en porcentaje; la expresión es:

$$THD + N = \frac{\sqrt{\sum_{i=1}^{\infty} v_{ih}^2 + V_{on}^2}}{V_o}100 \tag{2.48}$$

siendo v_{ih} las tensiones armónicas introducidas y no presentes en la señal de entrada y V_{on} es la tensión de ruido a la salida del operacional. Este parámetro es muy útil en las aplicaciones de audio.

El ruido no tiene el mismo ancho de banda aparente que la señal y se suele hablar del *ancho de banda equivalente de ruido (Equivalent Noise Bandwidth), ENB* [2.18], término que se usa para contabilizar la potencia de ruido que supera el ancho de banda real del circuito; por ejemplo, si tuviésemos un sistema de primer orden con ganancia 100 (40 dB) y un ancho de banda de 10 kHz, al superar el ancho de banda, la pendiente caerá con –20 dB/déc y una señal de 100 kHz sería amplificada en 20 dB, es decir, multiplicada por 10. Por ello, el ruido cuya frecuencia supere el ancho de banda será amplificado (menos que la señal pero amplificado a fin de cuentas) y se tendrá a la salida. El ENB tiene en cuenta este aspecto y, en función del orden del sistema, modifica el ancho de banda para el ruido (ver tabla 2.3).

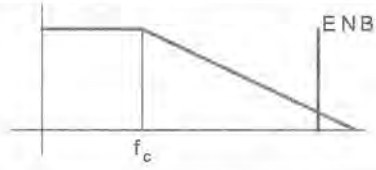
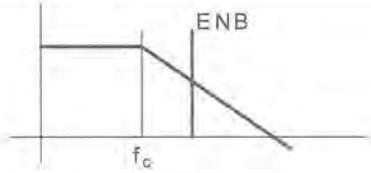
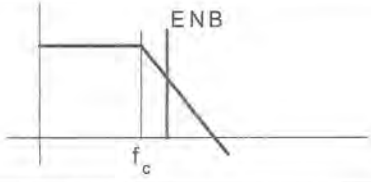
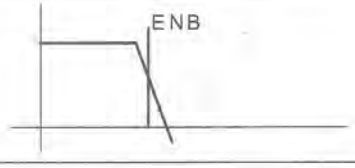
Orden del sistema	Diagrama de Bode (f_c es la frecuencia de corte)	ENB
1		$1{,}57 \cdot f_c$
2		$1{,}11 \cdot f_c$
3		$1{,}05 \cdot f_c$
4		$1{,}02 \cdot f_c$

Tabla 2.3. Ancho de banda equivalente para el ruido en sistemas de orden 1, 2, 3 y 4.

En las expresiones de cálculo del ruido, la integración debe hacerse entre los valores f_s y f_i pero estos no corresponderán a los márgenes de frecuencia que establezca el ancho de banda del circuito sino al valor modificado, es decir, $f_s - f_i$ = ENB y el cociente f_s/f_i puede aproximarse por la diferencia por lo que $f_s/f_i \approx$ ENB.

EJERCICIO 2.14

Sea un amplificador diferencial de ganancia 100 construido con resistencias R_1 = 1K y R_2 = 100K y que tiene un ancho de banda de 50 kHz. El operacional utilizado tiene una tensión de ruido de 18,0 nV/Hz$^{1/2}$ a 10 Hz y de 9,6 nV/Hz$^{1/2}$ para frecuencia de 1 kHz; de sus hojas de características encontramos que las co-

rrientes de ruido son de 0,32 pA/Hz$^{1/2}$ a 10 Hz y 0,12 pA/Hz$^{1/2}$ a 1 kHz. Calcular la relación S/N a la salida cuando se introduce una señal libre de ruido a la entrada con un valor de pico de 10,00 mV y el sistema está a una temperatura de 40°C.

SOLUCIÓN:

Es frecuente que el fabricante proporcione los valores de ruido en la forma aquí indicada en lugar de dar las curvas de densidades de tensión y de corriente de ruido en función de la frecuencia. No obstante, no es muy diferente ya que con estos datos es suficiente para hacer un cálculo, considerando que, a la mayor frecuencia, el ruido se debe sólo al ruido blanco, mientras que el valor de baja frecuencia se debe a la suma del blanco y el *flicker*, es decir, para las tensiones:

$$C = 9{,}6^2 = 92{,}16\text{nV}^2$$

$$C + \frac{k^2}{10\text{Hz}} = 18^2 \Rightarrow k^2 = 2318\text{nV}^2$$

$$f_{ncV} = \frac{2318}{92{,}16} = 25{,}16\text{Hz}$$

Para las corrientes, cálculos similares proporcionan los siguientes datos:

$$i_w^2 = 0{,}12^2 = 0{,}0144\text{pA}^2$$

$$i_w^2 + \frac{k^2}{10} = 0{,}32^2 \Rightarrow k^2 = 0{,}88\text{pA}^2$$

$$f_{ncI} = \frac{0{,}88}{0{,}0144} = 61{,}11\text{Hz}$$

El ancho de banda equivalente para el ruido será, según la tabla 2.3, 50·1,57 kHz = 78,5 kHz ya que el sistema será de primer orden como cualquier amplificador diferencial estable construido con un operacional.

Aplicando la expresión (2.47), el valor de la tensión eficaz de ruido a la salida del operacional teniendo en cuenta los conceptos de: ruido térmico en las resistencias, tensiones de ruido y corrientes de ruido del operacional, se obtiene que $V_{on} = 271{,}9\ \mu\text{V}$.

Como a la salida se producirá una señal de 1,000 V de pico, la relación señal-ruido será:

$$S/N = 20 \log \frac{1/\sqrt{2}}{271{,}9 \cdot 10^{-6}} = 68{,}3\ \text{dB}$$

Resumen

Como paso previo en el desarrollo de la lección se ha introducido la importancia de la amplificación para conseguir aportar potencia a una señal y hacerla, así, más insensible a las interferencias externas.

Después, ya de lleno en la presente lección, se ha dado un repaso a los diferentes aspectos que afectan al funcionamiento del operacional como dispositivo básico dentro de los sistemas de instrumentación. Para

ello, se han recorrido las diversas no-idealidades que presenta en su funcionamiento y se ha cuantificado lo que cada una de ellas aporta al error final.

El tema de la cuantificación es el que determina, en definitiva, si un determinado comportamiento puede ser admitido como válido o no y, por ello, permite saber qué no-idealidades afectan y cuáles pueden ser despreciadas.

El comportamiento del operacional se ha analizado desde los tres puntos de vista básicos: comportamiento estático, dinámico y generación de ruido. En cuanto al comportamiento estático se han cuantificado los efectos de diversos parámetros:

- Resistencias de entrada y salida.
- Ganancia en continua y ganancia de modo común.
- Tensiones de desviación y corrientes de polarización.
- Efectos de la tensión de alimentación.

En la tabla 2.4 se ha recogido la cuantificación de estos efectos sobre la configuración no inversora con operacional.

También se ha reiterado la importancia que tienen no sólo los parámetros en sí mismos, sino los cambios que vayan sufriendo a lo largo del tiempo como los que ocasione la temperatura (derivas) y la dificultad inherente a los ajustes.

Desde el punto de vista de su comportamiento dinámico se han tenido en cuenta diversos aspectos críticos de los amplificadores operacionales:

- Ancho de banda y producto ganancia-ancho de banda tanto para la ganancia diferencial como para la ganancia de modo común.
- El aspecto general de la ganancia diferencial como una función de la frecuencia y sus implicaciones *en la estabilidad de los circuitos.*
- El *slew-rate* como parámetro que define el ancho de banda en gran señal.

Finalmente, se ha tratado el comportamiento del amplificador en relación con el ruido, tanto el generado por él mismo como el producido por las resistencias que se usan en los circuitos de amplificación, y se ha cuantificado su efecto en el parámetro S/N.

Efectos de las no-idealidades estáticas sobre el amplificador operacional en configuración no inversora			
Parámetro		**Influencia**	**Cuándo hay que tenerlo en cuenta y otras consideraciones...**
$-V_{omax}$ $+V_{omax}$	Excursión de la tensión de salida (*output voltage swing*)	$-V_{o\,max} \le v_o \le +V_{o\,max}$	– **Siempre es una limitación** – **Importante si la tensión de alimentación es baja (equipos portátiles)** – En los operacionales *rail-to-rail* $V_{omáx}$ es casi igual a V_{CC}
A_d	Ganancia diferencial en lazo abierto (*Open loop gain*)	$v_o = \dfrac{A_d}{1 + A_d \dfrac{R_1}{R_1 + R_2}} v_i$	– **Con ganancias elevadas** (R_2 grande en relación con R_1) – **Con valores bajos de A_d** – **Cuando se tengan en cuenta la mayoría de los demás parámetros**
R_{id}	Resistencia de entrada (*Input resistance*)	$v_o = \dfrac{A_d}{1 + \dfrac{R_1 // R_2}{R_{id}} + A_d \dfrac{R_1}{R_1 + R_2}} v_i$	– **Con valores altos de resistencias** R_1 y R_2 – **Cuando la fuente de tensión de entrada tiene alta resistencia de salida**
R_o	Resistencia de salida (*Output resistance*)	$v_o = v_i \dfrac{A_d}{1 + \dfrac{R_o}{R_L} + \dfrac{R_o}{R_1 + R_2} + A_d \dfrac{R_1}{R_1 + R_2}}$	– Con niveles de carga en la salida elevados – Es un parámetro poco importante, en general
v_{io}	Tensión de desviación (*Input offset voltage*)	$v_o = v_{oideal} + v_{io} \dfrac{R_1 + R_2}{R_1}$ (expresión válida para la configuración inversora y para la diferencial)	– **Casi siempre que se tengan entradas con niveles de continua** – **Cuando la tensión de entrada es baja o muy baja** – Puede ser preciso realizar ajustes – Uno de los parámetros más importantes a la hora de seleccionar – Hay que cuidar la saturación por *offset* en casos de ganancia alta
I_B I_{io}	Corrientes de desviación y polarización (*Input bias and offset currents*)	$v_{ioTOTAL} = v_{io} + v_{dB}$ $v_{dB} = I_B \lvert R_{Th+} - R_{Th-} \rvert + I_{io} (R_{Th+} + R_{Th-})/2$ (se añade a la tensión de desviación)	– **Las mismas que en el caso de la tensión de desviación** – Aunque se puede eliminar el efecto de I_B, no siempre interesa – **Con valores altos de resistencias** R_1 y R_2
CMRR	Razón de rechazo del modo común (*Common mode rejection ratio*)	$v_o = \dfrac{A_d + A_c/2}{1 + (A_d + A_c/2) \dfrac{R_1}{R_1 + R_2}} v_i$	– Efecto sobre el valor de la ganancia – No tiene mucha importancia en el amplificador no inversor
A_d/A_c		$v_o = \dfrac{R_2}{R_1}(v_2 - v_1) + \dfrac{R_2}{R_1} \dfrac{v_2}{CMRR}$ (sólo para el amplificador diferencial)	– **Cuando hay valores elevados de la tensión de modo común** – **Cuando se pretende eliminar ruido de modo común** – ¡Cuidado con su comportamiento en frecuencia!
PSRR	Razón de rechazo a la alimentación (*Power supply rejection ratio*)	$PSRR = 20 \log \dfrac{\Delta V_{alimentación}}{\Delta v_o}$	– **En sistemas portátiles (alimentados a baterías)** – Puede ser conveniente incluir reguladores de tensión en la alimentación – Sólo tiene en cuenta variaciones lentas

Tabla 2.4

Problemas propuestos

2.1 Un operacional tiene una ganancia diferencial de 130 dB en continua y está siendo utilizado en un montaje no inversor de ganancia 10. Cuando en su entrada se sitúa una tensión de 0,1000 V, ¿qué error se produce en la salida?

2.2 Considérese un amplificador de ganancia −100 que tiene que amplificar una señal de 100,0 mV de amplitud y frecuencia 100 Hz. Considerando que la ganancia diferencial del operacional es de 100 dB en continua y tiene su primer polo en 10 Hz, determine el valor "real" de la tensión en su salida. ¿Hay mucha diferencia respecto a considerar la ganancia diferencial infinita?

2.3 Repita el problema anterior considerando la frecuencia de la señal como 10 kHz.

2.4 Un amplificador de entrada bipolar —como el 741— posee una resistencia de entrada mínima de 300K y típica, de 1M. Si la ganancia diferencial es de 106 dB en continua, ¿qué error se produce en la salida de un amplificador no inversor de ganancia 200 construido con él utilizando resistencias de 1K y 199K cuando recibe una señal de entrada de 10,000 mV?

2.5 Repita el mismo problema con un amplificador de entrada FET cuya resistencia de entrada es —como mínimo— de 10^6 MΩ.

2.6 ¿Cuál es la limitación de potencia de salida introducida por la resistencia de salida de un operacional de potencia cuya $R_o = 1\ \Omega$ cuando está alimentado entre +12 y −12V considerando que es capaz de alcanzar en la salida los valores de la tensión de alimentación?

2.7 Una etapa amplificadora tiene ganancia 100 y está realizada con un amplificador LM208 (ver hoja de características en el CD adjunto) cuya ganancia diferencial en continua es 110 dB y su tensión de desviación de entrada 3 mV. Calcular el error debido a la tensión de desviación que se introduce en su salida con una entrada de 10,00 mV.

2.8 Diseñe la red de compensación de *offset* para el problema anterior utilizando resistencias del 1% de tolerancia.

2.9 Supuesto ajustado el *offset* del problema 2.7 mediante la red diseñada en el problema 2.8, ¿cómo afecta la variación de la tensión de alimentación en el error de salida?

2.10 Compare los amplificadores AD549 y AD620 (ver hojas de características en el CD adjunto) respecto a sus tensiones de desviación. ¿Cómo se modificaría el resultado del problema 2.7 utilizando estos amplificadores?

2.11 Para el caso del problema 2.10, ¿sería conveniente utilizar una red de compensación de *offset*?

2.12 En el ejercicio 2.5 se plantea una solución para minimizar los efectos de la tensión de *offset* en la amplificación de una señal senoidal. ¿Es válido este sistema cuando la entrada es $v_i = 0{,}01000 + 0{,}0100\text{sen}(1000t)$? ¿Qué otra solución se usaría en este caso?

2.13 Un operacional *rail-to-rail* alimentado simétricamente a ±12 V está utilizado en un circuito inversor de ganancia −500 formado por dos resistencias de 1K y 500K y ha sido ajustado su *offset* total a una temperatura de trabajo de 20°C. Si está amplificando una señal senoidal de 20,00 mV de amplitud y 100 Hz de frecuencia y las máximas derivas son de 0,1 nA/°C y 0,01 mV/°C, ¿cuál es la forma de onda que podría encontrarse en el peor caso cuando el circuito trabaje a 60°C?

2.14 El operacional OP1177 (ver características en el CD adjunto) tiene baja corriente de polarización y sus características estáticas se pueden resumir en: 0,015 mV de tensión de *offset*, una corriente de polarización de 0,5 nA y una corriente de desviación de 0,2 nA. Si su ganancia diferencial en continua es de 120 dB (mínimo) determine el error cuando se está usando como amplificador inversor de ganancia −500 (R_1=1K y R_2=500K) para amplificar una señal de 1,0000 mV.

2.15 En el amplificador del problema anterior se compensa el error introducido por la corriente de polarización mediante una resistencia. Calcule dicha resistencia y el valor del nuevo error de salida. ¿Merece la pena efectuar esta compensación?

2.16 En el problema anterior, calcule el error máximo si la deriva térmica de la tensión de *offset* es de 0,002 mV/°C.

2.17 El amplificador OPA354 (ver características en el CD) se ha utilizado para realizar un amplificador diferencial de ganancia 100 mediante resistencias de 1K y 100K. Si su ganancia diferencial es de 94 dB y su CMRR de 66 dB, determine el error introducido para amplificar una señal que tiene un valor de tensión diferencial de 10,0 mV y una tensión de modo común de 3,00 V.

2.18 ¿Qué relación S/N tendría el amplificador del problema anterior si en la tensión de modo común fuese un ruido de 50 Hz de 3,00 V de valor eficaz y la tensión diferencial presentase un valor de 0,0200 V?

2.19 Un operacional LM748 (ver características en el CD) presenta una ganancia diferencial en continua de 106 dB presentando su primer polo en 50 Hz y su segundo polo en 1 MHz cuando está compensado con un condensador de 3 pF. ¿Cuál es la mínima ganancia que se puede permitir en un montaje no-inversor para que sea estable?

2.20 Indique cómo especificaría el producto GBW y el SR del operacional a utilizar para amplificar una señal que puede llegar a tener hasta 20 kHz y presenta un valor máximo de 50 mV de pico para ocupar todo el margen de entrada de un equipo que permite señales de hasta 5 V de pico.

2.21 Si para realizar el problema anterior sólo se dispone de un amplificador de suficiente SR pero la mitad del GBW necesario, ¿cómo podría diseñar el sistema?

2.22 Utilizando el operacional compensado OPA354 (ver características en CD) que posee un GBW de 100MHz diseñe un amplificador no inversor con ganancia 100. ¿Cuál es el ancho de banda obtenido?

2.23 Si el anterior diseño se alimenta entre 5 V y masa y teniendo en cuenta que el operacional es de tipo *rail-to-rail* (permite valores de la tensión de salida entre 0,2 y 4,8 V) y sabiendo que el SR es de 150 V/μs, ¿cuál es el máximo valor de la tensión de entrada para una frecuencia de 200 kHz?

2.24 En las mismas condiciones del problema anterior, ¿cuál es la máxima frecuencia de una señal de entrada que tenga un valor de pico a pico de 40 mV?

2.25 Dibuje la forma de onda de salida del circuito del problema anterior cuando la entrada es una señal cuadrada de 1,00 kHz y 40 mV de pico?

2.26 En las condiciones del problema anterior, ¿qué salida se produce con una señal de 10 MHz?

3

Circuitos amplificadores de uso en instrumentación

3.1 Amplificadores de instrumentación
3.2 Amplificadores aislados o de aislamiento
3.3 Amplificadores de transimpedancia y transconductancia
3.4 Otros tipos de amplificadores
3.5 Algunos criterios de selección de amplificadores operacionales para instrumentación
3.6 Otras consideraciones en el diseño de circuitos con amplificadores operacionales

En el capítulo anterior se trató el concepto de amplificación, se definieron los parámetros que caracterizan a un amplificador ideal y se expusieron las características de un amplificador real con la utilización del amplificador operacional como piedra angular para los circuitos amplificadores utilizados en instrumentación. En este capítulo estudiaremos los amplificadores específicos de uso en instrumentación y ofreceremos algunos criterios particulares de selección de los amplificadores operacionales.

3.1 Amplificadores de instrumentación

Un amplificador de instrumentación (AI) es un amplificador diferencial de tensión de precisión con un circuito optimizado para su trabajo en los ambientes más hostiles, caracterizados por grandes fluctuaciones de la temperatura e intenso ruido eléctrico. Además, estos amplificadores especiales deben ser capaces de trabajar con sensores de resistencia interna apreciable y no simétrica, sobre los que el ruido eléctrico inducido y/o conducido tiene una gran influencia y, por si esto fuera poco, que ofrecen señales eléctricas muy débiles [3.3].

3.1.1 PARTICULARIDADES DEL AMPLIFICADOR DE INSTRUMENTACIÓN

Para ser efectivo, un amplificador de instrumentación debe ser capaz de amplificar señales del orden de microvoltios y a la vez rechazar tensiones de modo común del orden de voltios. Esto presupone que el amplificador de instrumentación tenga un elevado rechazo al modo común CMRR (del orden de 80 ó 100dB) a la vez que una gran ganancia diferencial A_d. A continuación, expondremos las configuraciones más utilizadas como amplificadores de instrumentación.

Amplificador diferencial

En la figura 3.1 se muestra el circuito típico de un tipo de amplificador de instrumentación conocido como amplificador diferencial o simple restador.

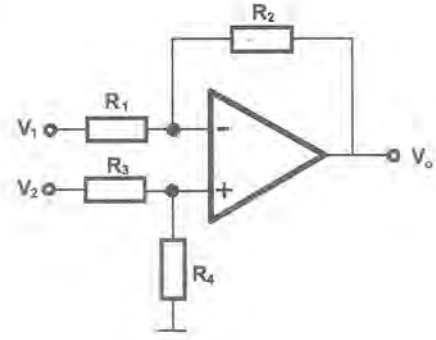


Figura 3.1. Amplificador diferencial.

En este circuito puede demostrarse, si se considera el AO utilizado como ideal, que la salida v_o puede representarse por la expresión:

$$v_o = \frac{R_2}{R_1 + R_2} \cdot \frac{R_3 + R_4}{R_3} v_2 - \frac{R_4}{R_3} v_1 \tag{3.1}$$

y, si hacemos que $R_2 = R_4$ y $R_1 = R_3$, entonces (3.1) queda como:

$$v_o = \frac{R_4}{R_3}(v_2 - v_1) \tag{3.2}$$

Si $v_1 = v_2 = v_{ic}$ en (3.1), tendremos la salida para una entrada de modo común, que será:

$$v_{oc} = \left(\frac{R_2}{R_1 + R_2} \cdot \frac{R_3 + R_4}{R_3} - \frac{R_4}{R_3} \right) v_{ic} \tag{3.3}$$

Como ya se demostró en la ecuación (2.30), aun cuando se consideren todas las resistencias iguales entre sí e iguales a R, la salida v_o no sólo dependerá de la diferencia entre v_2 y v_1, sino también de la razón de rechazo al modo común (CMRR) muy dependiente del balance real entre estas resistencias.

En el amplificador diferencial, las resistencias de entrada vistas desde cada entrada del amplificador a masa son bajas y desiguales (la vista por v_1 es aproximadamente R_1 mientras que la vista por v_2 es $R_3 + R_4$).

Por otro lado, para cambiar la ganancia hay que cambiar dos resistencias[1] y, con una muy pequeña desviación entre ellas, CMRR desciende drásticamente. Si la resistencia interna de la fuente de señal R_S no es despreciable y/o equilibrada respecto a la masa del circuito, entonces R_S (o su desequilibrio respecto a masa) contribuye a hacer desiguales las resistencias del amplificador y a la degradación tanto de la ganancia diferencial A_d como del rechazo al modo común del circuito.

A pesar de estas limitaciones los operacionales constituyen el primer escalón del desarrollo de los denominados amplificadores de instrumentación. En el CD se ofrece la hoja de características del INA105, fabricado por Burr-Brown, caracterizado por la precisión de sus resistencias ajustadas mediante láser. La figura 3.2 muestra su esquema simplificado.

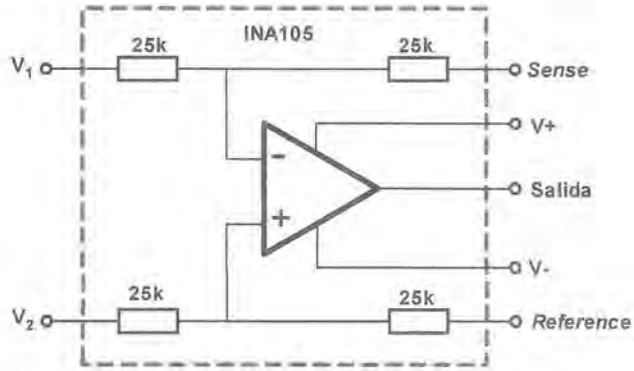


Figura 3.2. Estructura interna del INA105.

El INA105 es un amplificador diferencial de ganancia unitaria (resistencias iguales) en el que el acceso de forma independiente a sus resistencias posibilita su uso en un gran número de configuraciones amplificadoras, además de la diferencial, tales como fuentes de corriente controladas por tensión, amplificadores

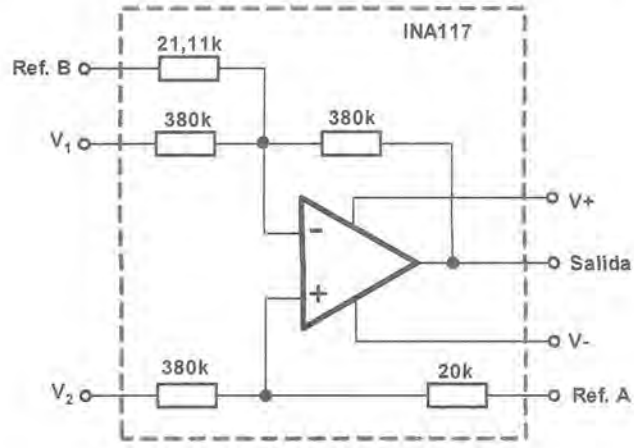


Figura 3.3. Estructura interna del INA117.

[1] El amplificador AD628 de Analog Devices es de ganancia programable por resistencias externas que cambian la ganancia de una etapa amplificadora no inversora de salida. La ganancia de la etapa diferencial de entrada es fija e igual a 0,1. Su CMRR es de 75 dB para G_{TOTAL}=1 desde continua hasta 500 Hz.

inversores de ganancia 1, no-inversores de ganancia 1, 2 ó 0,5 sin el uso de componentes externos, sumadores, convertidores I/V y V/I para bucles de corriente de 4÷20 mA, así como para filtros activos lo que pone de manifiesto el gran campo de aplicación en instrumentación de este dispositivo. Existen también otros amplificadores diferenciales monolíticos capaces de permitir en su entrada muy altos niveles de tensión de modo común, como el INA 117, cuyos datos del fabricante se ofrecen en el CD. La figura 3.3 presenta la estructura interna de este amplificador.

La incorporación al circuito de otras resistencias desiguales para compensar la desigualdad de las resistencias de entrada, hace posible además conformar un divisor resistivo de tensión en cada entrada externa hasta las entradas + y – del operacional que permite aumentar significativamente el rechazo al modo común del circuito [3.8]. Si en el circuito de la figura 3.3 se interconectan los terminales Ref A y Ref B a la masa de las fuentes de alimentación V^+ y V^-, entonces el circuito será capaz de trabajar con tensiones de modo común de hasta un máximo de 200 V. La principal limitación del amplificador diferencial es su poca capacidad de ofrecer una elevada ganancia diferencial A_d ajustable fácilmente junto con una muy pequeña ganancia al modo común, A_c.

Amplificador de instrumentación de dos operacionales

El amplificador de instrumentación de dos operacionales es una configuración que presenta resistencias de entrada altas e iguales lo que permite que la fuente de señal pueda tener una impedancia interna alta y/o desequilibrada, mejorando en este sentido la limitación del amplificador diferencial. Su esquema general se presenta en la figura 3.4a.

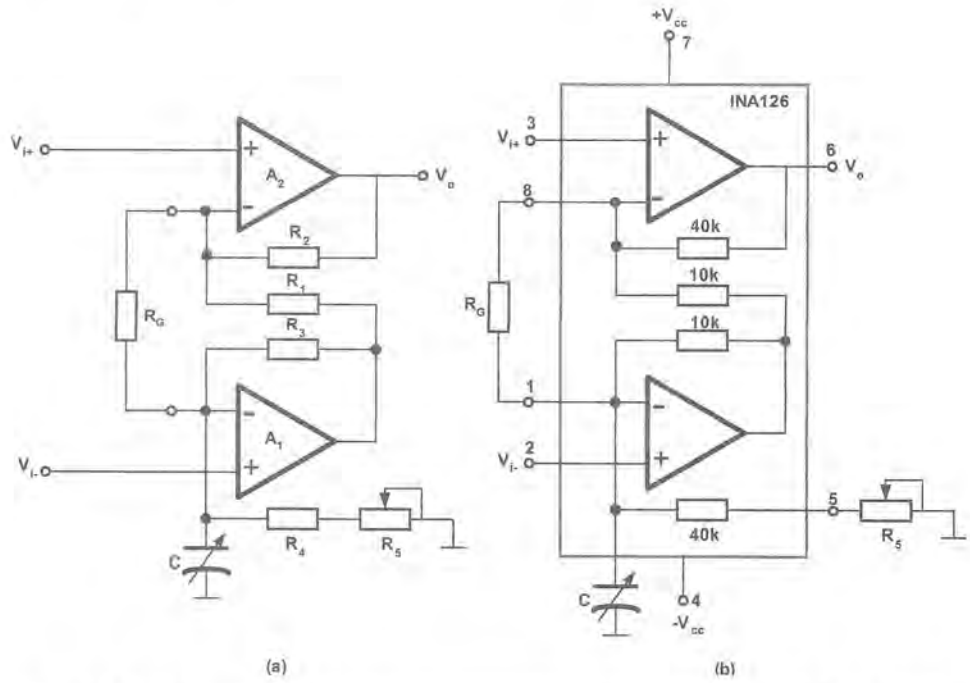


Figura 3.4. a) Amplificador de dos operacionales; b) Estructura interna del INA126.

Mediante un análisis similar al circuito anterior y considerando ideales los operacionales A_1 y A_2, puede demostrarse que si $R_2/R_1 = R_4/R_3$, entonces:

$$v_o = \left(1 + \frac{R_2}{R_1} + 2\frac{R_2}{R_G}\right)(v_{i+} - v_{i-}) \tag{3.4}$$

La ganancia diferencial A_d puede seleccionarse variando sólo R_G (resistencia externa si el circuito es un circuito integrado). En este amplificador no se puede lograr $A_d = 1$ y aun la razón de rechazo al modo común (CMRR) depende de la paridad entre las resistencias. El potenciómetro R_5 sirve de ajuste del rechazo al modo común en continua, mientras que C ajusta este en alterna [3.2]. La principal desventaja de este circuito es que su margen de entrada para la tensión de modo común, v_c es una función de la ganancia. De la figura 3.4 podemos observar que A_1 puede saturarse con una tensión de modo común relativamente baja, pues le ofrece a esta señal una ganancia dependiente de $(R_3 + R_4 + R_5)/(R_4 + R_5)$. Si $R_3 > (R_4 + R_5)$, ocurrirá la saturación de A_1 para altos valores de v_c; si, por el contrario $R_3 < (R_4 + R_5)$ no se pueden obtener en el circuito pequeñas ganancias y además los dos amplificadores A_1 y A_2 trabajarán con diferentes ganancias en bucle cerrado (diferentes anchos de banda) y esto trae como consecuencia un detrimento de la CMRR para alterna de no utilizarse C en el circuito. Esto hace que el circuito se vea limitado sólo para aplicaciones de pequeñas tensiones de modo común y, en general, en las que se cumpla:

$$(R_3 + R_4 + R_5)/(R_4 + R_5) \leq 4$$

La figura 3.4b presenta el circuito interno del INA126, circuito integrado en el que se cumple:

$$\frac{R_3 + R_4}{R_4} = 1{,}25$$

Para este circuito, se tiene que

$$1 + \frac{R_2}{R_1} + 2\frac{R_2}{R_G} = G \Rightarrow G = 5 + \frac{80\ k\Omega}{R_G}$$

y, por consiguiente, la tensión de salida de A_1 será:

$$v_{oA_1} = 1{,}25\, v_{i-} - \frac{10\ k\Omega}{R_G}(v_{i+} - v_{i-})$$

En cualquier modo de trabajo del circuito las tensiones de entrada deben ser tales que V_{oA1} no alcance la saturación. En el caso particular de este circuito integrado, la saturación está a 0,7 V del valor de V_{cc}. Las hojas de características del fabricante del INA126 se incluyen en el CD adjunto.

EJERCICIO 3.1

Diseñe un amplificador de instrumentación utilizando el INA126, alimentado a ±15 V que entregue una tensión de 0 a 5 V de pico a una carga resistiva pura de 1 kΩ cuando es excitado por una fuente de señal senoidal de 0 a 50 mV de pico y 500 Hz de frecuencia. ¿Qué valor de tensión de modo común permite el circuito a su entrada sin producir distorsión en la salida de A_1?

SOLUCIÓN:

La ganancia del circuito será G = 5 V/50 mV = 100. De la ecuación del fabricante para este circuito se tiene que 100 = 5 + 80 kΩ/R_G con lo que R_G resulta ser de 842 Ω. De la hoja de características del fabricante, para G = 100 obtenemos que la frecuencia de corte del amplificador se encuentra aproximadamente en 5 kHz con lo que no existirá distorsión de frecuencia para esa ganancia en la salida del amplificador.

Para el caso de este problema, la tensión de salida de la etapa A_1 será:

$$v_{oA_1} = 1{,}25\, v_{i-} - 11{,}88 \quad (v_{i+} - v_{i-})$$

Si consideramos que la tensión a la que A_1 se satura se encuentra en 15 V – 0,7 V = 14,3 V, en ningún caso v_{oA1} debe alcanzar esta tensión.

Para señal nula, considerando una entrada de modo común $v_c = v_i$, el valor mayor de v_c será:

$$v_c = \frac{14{,}3}{1{,}25} = 11{,}44 \text{ V}$$

Pero, para máxima señal (50mV) y en el semiciclo en que $v_{i+} - v_{i-}$ sea negativo y máximo, se cumplirá:

$$v_{oA_1} = 1{,}25 \left(v_c + 25 \cdot 10^{-3}\right) - 50 \cdot 10^{-3} \cdot 11{,}88 \leq 14{,}3 \text{ V}$$

Y, de la condición "igual a", se obtiene que $v_{c\,máx} = 11{,}89$ V.

Por lo que la tensión de modo común máxima para que este circuito no ofrezca distorsión de amplitud a la salida de A_1 debida a su saturación será de 11,9 V.

Amplificador de instrumentación de tres operacionales

Un circuito que integra las ventajas de los anteriores y reduce en buena medida sus desventajas es el conocido como amplificador de instrumentación de tres operacionales, o simplemente **amplificador de instrumentación**. La figura 3.5 presenta su circuito.

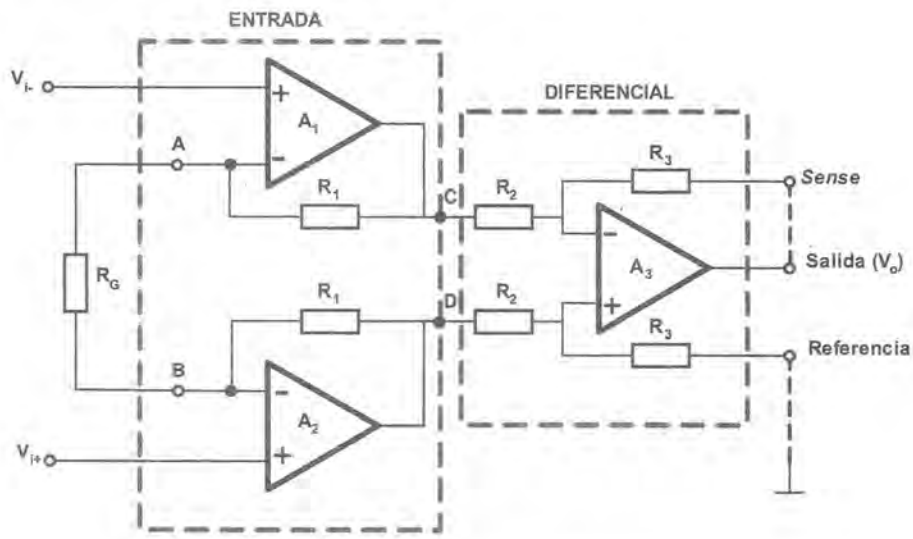


Figura 3.5. Amplificador de instrumentación.

En él se advierten dos etapas: la de **entrada** y la **diferencial**. Si en la etapa de salida se conecta el terminal *Sense*[2] al de Salida y el terminal Referencia a masa, tenemos un amplificador diferencial. La primera etapa cumple la función esencial de presentar impedancias de entrada altas e idénticas, a la vez que hace posible que el amplificador diferencial tenga como fuente de señal las salidas del mismo, caracterizadas por unas impedancias muy pequeñas e iguales, con independencia de que la fuente de señal externa del circuito tenga su resistencia R_S equilibrada o no respecto a la masa. En esta etapa, si los operacionales A_1 y A_2 son idénticos y también lo son las resistencias R_1, la ganancia puede llegar a ser mucho mayor que 1, dependiendo sólo de los valores de R_1 y R_G. Fácilmente puede demostrarse por superposición [3.2], considerando ideales los operacionales, que, para $v_{i+} = 0$ se tiene:

$$v_C = \frac{R_1 + R_G}{R_G} v_{i-} \quad y \quad v_D = \frac{R_1}{R_G} v_{i-} \tag{3.5}$$

Y, cuando $v_{i-} = 0$ se tiene que:

$$v_C = \frac{R_1}{R_G} v_{i+} \quad y \quad v_D = \frac{R_1 + R_G}{R_G} v_{i-} \tag{3.6}$$

Resultando:

$$v_C = \frac{R_1 + R_G}{R_G} v_{i-} + \frac{R_1}{R_G} v_{i+}$$

$$v_D = \frac{R_1}{R_G} v_{i-} + \frac{R_1 + R_G}{R_G} v_{i+}$$

La salida del amplificador diferencial final será, entonces:

$$v_o = \frac{R_3}{R_2} \cdot \left(1 + \frac{2R_1}{R_G}\right)\left(v_{i+} - v_{i-}\right) \tag{3.7}$$

En este circuito la exactitud de la ganancia diferencial y el rechazo al modo común dependen de la igualdad entre sí de las resistencias R_2 y de las R_3. Incluso si existe alguna desigualdad entre las resistencias R_1, esta no afecta al rechazo al modo común. De esta forma, CMRR puede incrementarse en proporción directa a la ganancia diferencial A_d. Además, si la etapa de entrada es simétrica, por R_G no fluye corriente alguna debida a la tensión de modo común por lo que la ganancia que ofrece esta etapa al modo común es la unidad. Como quiera que cuanto antes se aporte ganancia a una señal, mejor relación S/N tendremos, el circuito integrado que constituye el amplificador de instrumentación produce toda la amplificación en la primera etapa y, generalmente la etapa diferencial se construye con $R_2 = R_3$, por lo que de la ecuación (3.7) se obtiene A_d:

$$A_d = 1 + \frac{2R_1}{R_G} \tag{3.8}$$

Si se desease que $A_d = 1$, entonces R_G debe ser infinita. El valor máximo de la ganancia diferencial A_d en los circuitos integrados monolíticos comerciales puede llegar a alcanzar hasta 1000 y R_1 es del orden de algunos kilohmios.

---

[2] El terminal *Sense* es utilizado para que al conectar una carga remota las condiciones de tensión de salida v_o y ganancia del circuito A_d se cumplan ***en la carga***, con independencia de la resistencia de los conductores que unan la carga con el circuito.

Con el propósito de hacer posible valores de ganancia no muy comunes el amplificador de instrumentación AD620 (que es una variante mejorada del amplificador de tres operacionales cuyas hojas de datos se encuentran en el CD adjunto), utiliza dos resistencias internas de muy alta precisión que junto con la resistencia R_G externa determinan la ganancia del circuito. Esto permite conseguir valores muy estables de la ganancia empleando resistencias de alta precisión para R_G.

EJERCICIO 3.2

Utilizando el amplificador de instrumentación AD620 calcule los valores de R_G necesarios para lograr ganancias de 1, 2, 5, 20 y 200.

NOTA: La expresión que proporciona el fabricante para calcular la resistencia externa R_G a utilizar es $R_G = 49{,}4\ k\Omega/(G-1)$.

SOLUCIÓN:

Aplicando la expresión del fabricante se tiene que para conseguir ganancia 1 habrá que situar una resistencia infinita, es decir, los terminales 1 y 8 del circuito quedarán no conectados.

De la misma forma, para conseguir ganancia 2, se precisará una resistencia de 49,4 kΩ y el valor estándar más próximo en la serie del 1% de tolerancia es 49K9, lo que deja la ganancia final obtenida en G = 1,99. Si recurrimos a la serie estándar del 0,1% de tolerancia (considerablemente más cara que la anterior), el valor más próximo es 49K3 lo que deja la ganancia en G = 2,002.

Para conseguir ganancia 5, se precisará una resistencia de 12,35 kΩ y el valor estándar más próximo tanto en la serie del 1% de tolerancia como en la del 0,1% es 12K4, lo que deja la ganancia final obtenida en G = 4,984.

Si la ganancia debe ser 20, se precisará una resistencia de 2,6 kΩ. El valor estándar más próximo tanto en la serie del 1% de tolerancia como en la del 0,1% es 2K61, lo que deja la ganancia final obtenida en G = 19,93.

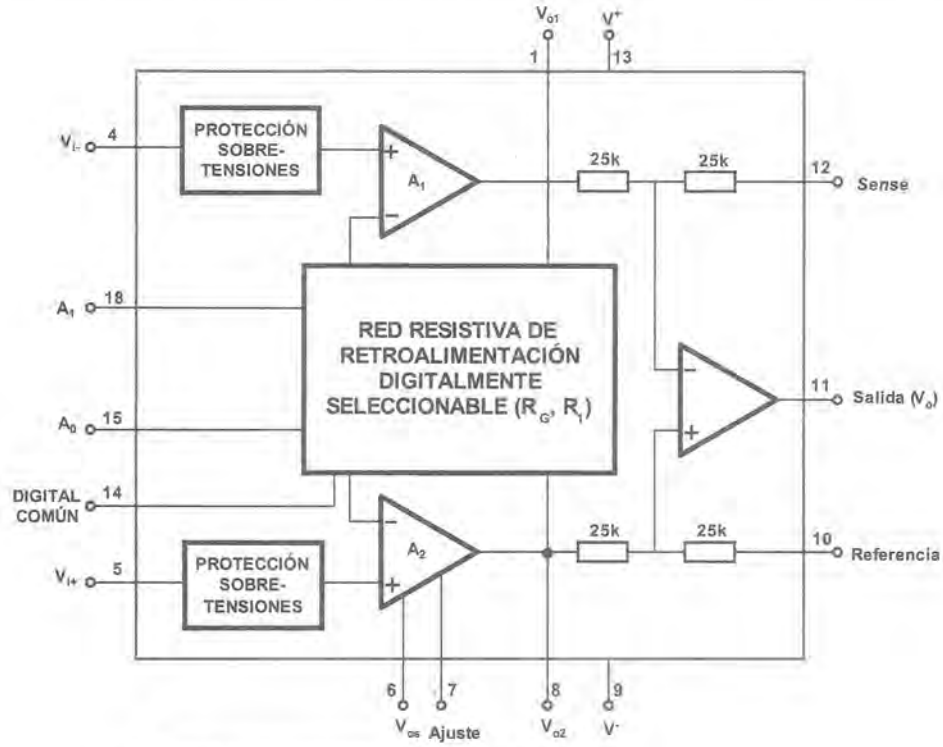
Finalmente, para ganancia de 200, se precisará una resistencia de 248 Ω. El valor estándar más próximo tanto en la serie del 1% de tolerancia como en la del 0,1% es 249, con lo que la ganancia final será G = 199,4.

Es posible la obtención de ciertos valores discretos de la ganancia diferencial A_d sin utilizar ninguna resistencia externa en los denominados **amplificadores de instrumentación de ganancia programable por pin** (*pin-programmable*). Estos amplificadores poseen internamente un conjunto de resistencias R_G de muy alta precisión que garantizan determinados valores muy exactos y estables de A_d, cuyos valores generalmente se disponen en potencias de diez o de dos. La elección de la ganancia se realiza simplemente uniendo pines del circuito. En el CD adjunto se encuentran las hojas de características del AD524, capaz de conseguir ganancias de 1, 10, 100 y 1000.

Otra versión de estos amplificadores es la conocida como **amplificador de instrumentación de ganancia programable por software** (*Software Programmable Gain Instrumentation Amplifier*), que permiten la selección por software de diferentes valores de la ganancia. Como en los anteriores, los valores discretos de la ganancia en estos amplificadores suelen estar en potencias de 10 (1, 10, 100, 1000) o en potencias de 2, es decir, 1, 2, 4, 8, 16, 32... A la cantidad de valores discretos se le llama **pasos** [3.6].

La figura 3.6 ilustra la arquitectura interna de los amplificadores programables PGA204 y 205. El PGA204 permite 4 pasos y sus ganancias están en potencias de 10 (1, 10, 100 y 1000), mientras que el PGA205, también de 4 pasos, tiene sus ganancias en potencias de 2 (1, 2, 4 y 16). También puede diseñarse

un amplificador programable partiendo de un amplificador de instrumentación y un circuito selector analógico que conmute diversas resistencias R_G con diferentes valores de resistencias en función del código de selección, posibilitando tener las ganancias deseadas asociadas a los diferentes códigos.



GANANCIA		SELECCIÓN	
PGA204	PGA205	A_1	A_0
1	1	0	0
10	2	0	1
100	4	1	0
1000	8	1	1

Figura 3.6. Estructura interna de los amplificadores programables PGA204 y 205.

La figura 3.7 presenta un circuito con un multiplexor analógico diferencial que conecta sus entradas 1, 2, 3 ó 4 a la salida en función de un código binario A_0A_1 que puede valer 00, 01, 10 ó 11.

Puede observarse que, en este caso, el código 00 hace que A_d sea igual a uno y, en los restantes casos, el valor de A_d estará en dependencia de los valores de R_2, R_3 y R_4. La utilización de este circuito de selección introduce un error debido a su resistencia en conducción (R_{ON}). Esta resistencia es idealmente nula pero en la práctica no es así, influyendo de forma desfavorable para ganancias muy grandes donde R_G toma un valor pequeño, comparable a la resistencia de conducción.

Para conseguir el valor exacto de la ganancia deberá considerarse como valor real de la resistencia R_G el correspondiente a la suma de la resistencia que pongamos externamente y la resistencia total del circuito de selección en conducción[3].

[3] Si se utilizan como selectores multiplexores analógicos de tecnología CMOS, la resistencia que se puede encontrar en conducción es de algunas decenas de ohmios. Estos valores pueden tener una gran importancia en aplicaciones con ganancias altas, donde R_G es del orden de una o dos centenas de ohmios (ver ejercicio 3.2).

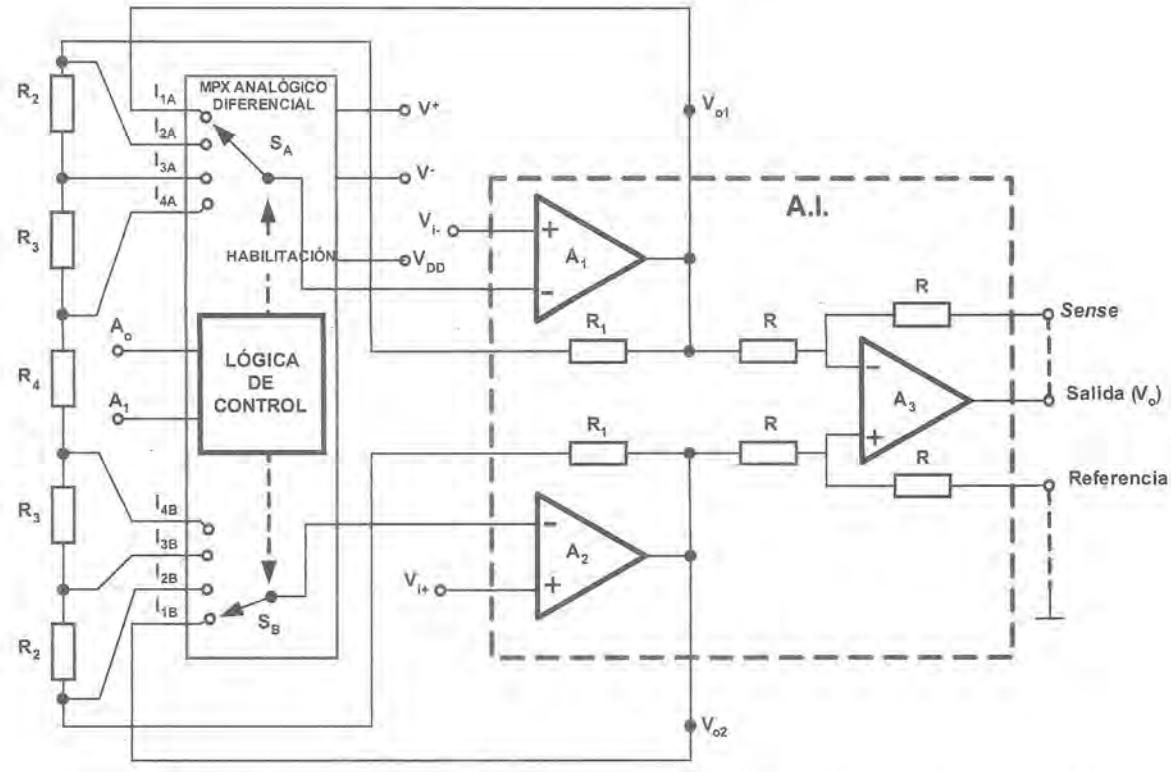


Figura 3.7. Amplificador de instrumentación de ganancia programable por software con A.I. y multiplexor analógico.

Estos amplificadores se utilizan fundamentalmente en sistemas de medida de sensibilidad variable y en sistemas multicanal de multiplexado en tiempo donde varios sensores (generalmente de la misma naturaleza) son conectados mediante un multiplexor analógico en forma consecutiva a la entrada del amplificador, pudiendo este ofrecer diferentes ganancias en cada caso (ver capítulo 22).

EJERCICIO 3.3

En la figura 3.8 se muestra un sistema que pretende llevar 4 señales diferentes (cada una de ellas procedente de un sensor) a un sistema de adquisición de datos. Los diferentes márgenes de variación de cada una de las señales fuerzan a usar ganancias diferentes en los amplificadores. Suponiendo que se emplea un amplificador de instrumentación AD620 diseñe un sistema de ganancia programable utilizando el multiplexor analógico 4052.

NOTAS:

- *La expresión que proporciona el fabricante para calcular la resistencia externa es:*

$$R_G = 49{,}4k\Omega/(G\text{-}1).$$

- *El fabricante del 4052 proporciona como dato una resistencia en conducción máxima que oscila entre 86 Ω para temperaturas entre -55 y 25°C y 120 Ω cuando la temperatura es la extrema de trabajo de 125°C.*

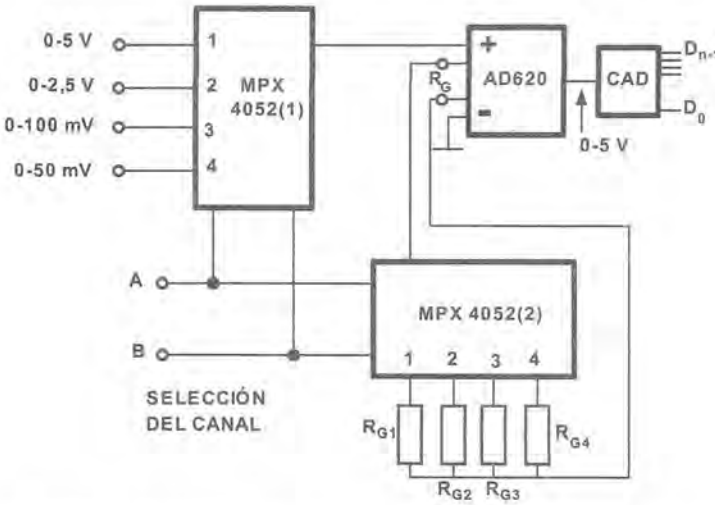


Figura 3.8. Sistema de adquisición de datos. Ejercicio 3.3.

SOLUCIÓN:

Como quiera que el amplificador debe entregar una tensión comprendida entre 0 y 5V, las ganancias correspondientes a cada canal son:

CANAL 1: ganancia unitaria (5/5).
CANAL 2: ganancia 2 (5/2,5).
CANAL 3: ganancia 50 (5/0,1).
CANAL 4: ganancia 100 (5/0,05).

En estas condiciones, la selección del canal implica la conexión de una resistencia específica entre los terminales del AD620 ya que las señales lógicas A y B controlan ambos multiplexores. De esta forma, la resistencia total que debería verse sería —aplicando la expresión de la ganancia proporcionada por el fabricante— para cada uno de los canales:

CANAL 1: la resistencia será infinita.
CANAL 2: la resistencia será de 49,4 kΩ.
CANAL 3: el valor total de la resistencia deberá ser 1,008 kΩ.
CANAL 4: se obtendrá una resistencia de 499 Ω.

Como quiera que hay que tener en cuenta la resistencia en serie que impone el 4052 en conducción, los valores de R_{G1}, R_{G2}, R_{G3} y R_{G4} resultan ser ∞, 49,3 kΩ, 922 Ω y 413 Ω, respectivamente.

Si acudimos a la serie del 1% de valores estándar de resistencias (no tiene sentido acudir a una serie más precisa ya que el propio multiplexor introduce una incertidumbre más elevada), las más próximas serán R_{G2} = 49K3, R_{G3} = 919 Ω y R_{G4} = 412 Ω.

El lector puede recalcular las ganancias en función de estos valores y obtendrá alguna diferencia respecto de los valores previstos para los canales 3 y 4. No obstante, debe tener en cuenta que la dispersión de valores introducirá errores adicionales en todos los canales (excepto en el 1 en que la aproximación al infinito por el circuito abierto es bastante aceptable).

3.1.2 PARÁMETROS QUE CARACTERIZAN A UN AMPLIFICADOR DE INSTRUMENTACIÓN

Los parámetros fundamentales que caracterizan a los amplificadores de instrumentación son parecidos a los que se definieron en el capítulo 2: parámetros estáticos (comportamiento en continua) y parámetros dinámicos (comportamiento en alterna). A continuación, se indican los más importantes:

1. **Ganancia (*Gain*):** expresa la función de transferencia lineal del amplificador, es decir, el factor de amplificación de tensión al modo diferencial A_d. En los circuitos integrados monolíticos, por lo general, el fabricante ofrece una expresión analítica para la misma, que denotan como G y que, obviamente es una función de la frecuencia. Ya hemos visto algún ejemplo de este tipo de expresiones en los ejercicios 2.2 y 2.3 y ésa será la tónica habitual, como en el caso del AD524 en que la expresión es $G = 1 + 40000/R_G$.

2. **Margen de ganancia (*gain range*):** es el margen de valores en que puede estar contenida la ganancia diferencial del amplificador de instrumentación e indica la capacidad de amplificación del dispositivo. En los casos de amplificadores de ganancia programable por pin o por software, además de sus valores extremos, se indica la cantidad de pasos. Por ejemplo, se indicará que G varía entre 1 y 1000, y que el valor concreto se obtiene variando R_G.

   En un amplificador programable, se indicará que G está comprendida entre 1 y 1024 en 11 pasos lo que significa que la ganancia se puede fijar por software a cualquiera de estos valores: 1, 2, 4, 8, 16, 32, 64, 128, 256, 512 ó 1024.

3. **Error de la ganancia (*gain error*):** es la desviación de la ganancia real respecto a la ideal, para una R_G constante. Se suele expresar en tanto por ciento respecto al valor de ganancia mayor del margen y/o se puede especificar para cada posible ganancia (en el caso de los programables).

4. **Deriva de ganancia (*gain drift*):** es la variación en la ganancia diferencial producida por cambios en la temperatura, expresada generalmente en ppm/°C. Por ejemplo, una deriva de ±5 ppm/°C implica que si la temperatura del circuito varía 50°C, el cambio en la ganancia es de 250 ppm o, lo que es lo mismo, un valor absoluto de 0,00025G.

5. **Tensión de desviación (*offset voltage*):** es el mismo concepto que se manejó en el capítulo anterior, es decir, la magnitud de la tensión que aparece en la salida del amplificador cuando sus entradas están unidas y conectadas al punto de referencia de la alimentación (0V). En el caso de los amplificadores de instrumentación puede estar **referida a la entrada**, en cuyo caso se denota con las siglas RTI (*Referred To Input*) o la salida (se indica como RTO o *Referred To Output*). En el segundo caso, el valor es directamente la tensión de desviación de salida mientras que, en el primero, para calcular su magnitud en la salida, es necesario afectar su valor por la ganancia diferencial con que trabaja el amplificador. Una gran parte de los fabricantes de amplificadores de instrumentación ofrecen este dato con la indicación RTI. Por ejemplo, un amplificador que tenga una tensión de desviación de 20 μV (RTI) producirá una tensión de desviación de 2 mV cuando trabaje con ganancia 100.

   Sin embargo, recordemos que la mayoría de los amplificadores de instrumentación son circuitos de dos etapas; por lo general, una de ellas de ganancia variable (la de entrada) y otra de ganancia fija (la de salida), por lo que la tensión de desviación tiene dos componentes introducidas por cada una de las dos etapas:

   Si v_{io1} es el *offset* de la primera etapa y v_{io2} el de la segunda, se tiene que:

$$V_{io} = V_{io1} + \frac{V_{io2}}{G} \tag{3.9}$$

Por ejemplo, un amplificador puede proporcionar su tensión de desviación como ±10 ±20/G μV (RTI). Proporcionar el dato referido a la entrada tiene como finalidad conocer los extremos de ajuste de esta tensión, tarea que se lleva a cabo sobre la entrada. En algunos amplificadores existe la posibilidad de compensar (mediante circuito externo generalmente asociado a la primera etapa) la tensión de desviación inicial para una ganancia dada por lo que la deriva de la tensión de desviación entonces pasa a tener un alto protagonismo. En los amplificadores de ganancia programable por software se optimiza el circuito para muy bajos niveles de tensión de desviación; si es necesario se ajusta el *offset* inicial a cero para la mayor ganancia con las que trabajará el circuito.

6. **Deriva térmica de la tensión de desviación (*offset voltage drift*):** es la variación de la tensión de desviación del amplificador producida por los cambios de la temperatura en el dispositivo ($\Delta v_{io}/\Delta T$). Este valor se suele expresar también referido a la entrada (RTI).

EJERCICIO 3.4

a) Calcule la tensión de desviación de un amplificador de instrumentación de ganancia diferencial G = 100 y margen de la señal de entrada de 10 mV a 65°C conociendo que su tensión de desviación es de 20 μV (RTI) a 25°C y su deriva térmica es de 2 μV/°C (RTI).

b) Calcule el error debido a la deriva de tensión a ambas temperaturas.

c) ¿Qué tensión representa este error en la salida?

SOLUCIÓN:

a) La tensión de desviación referida a la entrada a 65°C será:

$$v_{io\,65°C}(\text{RTI}) = v_{io\,25°C}(\text{RTI}) + {}^{\Delta v_{io}}\!/_{\Delta T} \cdot (65-25)°\text{C} \Rightarrow v_{io\,65°C}(\text{RTI}) = 100\ \mu\text{V}$$

b) Teniendo en cuenta que la señal de entrada es de 10mV, el error relativo a 25°C será de 20 μV/10 mV, es decir, un 0,2%.

A la temperatura de 65°C, el error será mayor según lo calculado en el apartado a), obteniéndose 100 μV/10 mV o, lo que es lo mismo, un 1%.

c) A la salida, la tensión de *offset* se obtendrá sin más que multiplicar por la ganancia del dispositivo con lo que aparecerán desviaciones de 2 mV a 25°C y de 10 mV a 65°C.

---

7. **Corrientes de polarización de entrada y de desviación (*bias and offset input currents*):** corresponde al mismo concepto ya estudiado para los operacionales y su causa es la misma puesto que este tipo de amplificadores tendrá como bloques integrantes amplificadores operacionales cuyas entradas son las del amplificador de instrumentación. Así, la corriente de polarización de entrada, I_B es la que circula por cualquiera de las entradas del amplificador mientras que la corriente de desviación I_{io} es la diferencia entre las dos corrientes de polarización.

En las aplicaciones con sensores sin ningún terminal referido a masa, resulta necesario colocar resistencias desde cada entrada del amplificador a masa para que por ellas circule las corrientes de polarización, evitando la saturación de los amplificadores de entrada. Estas resistencias deberán ser iguales entre sí para que la corriente de polarización que circule por ellas produzca idénticas caídas de tensión y no contribuya a incrementar la tensión de desviación en la entrada del dispositivo.

Si la fuente de señal que conectamos a la entrada tiene resistencias a masa, las corrientes de polarización que circulan por esas resistencias generarán tensiones de desviación si estas resistencias

no son iguales entre sí (figura 3.9a). La tensión generada se añadirá a la tensión de desviación propia del amplificador para obtener una tensión de desviación total que determinará el error del sistema.

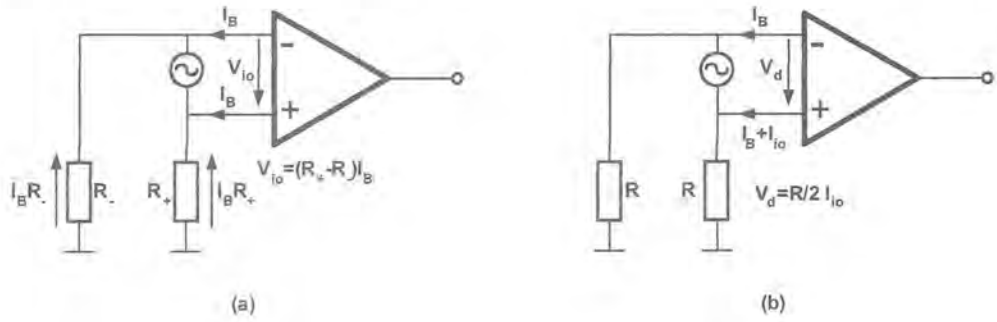


Figura 3.9. Circuito de la fuente de señal: a) efecto de las corrientes de polarización; b) efecto de la corriente de desviación.

Aun en el caso en que ambas resistencias sean iguales, la corriente de desviación ocasionará una tensión de desviación que será mayor cuanto mayor sea el valor óhmico de tales resistencias[4] (figura 3.9b).

8. **Derivas térmicas de las corrientes de polarización y desviación (*bias and offset input currents drift*):** estos parámetros tienen igual interpretación que la deriva térmica de la tensión de desviación del amplificador y corresponden con el concepto ya expresado para los operacionales.

EJERCICIO 3.5

Un amplificador de instrumentación se utiliza para amplificar la señal de tensión que ofrece un puente de Wheatstone con un sensor resistivo R_X. El circuito equivalente del puente es una fuente de señal v_s con dos resistencias R_{S1} y R_{S2} que durante la operación del puente tendrán un desequilibrio máximo de 500 Ω.

a) Calcule el error debido a la circulación de las corrientes de polarización por dichas resistencias si el margen de la señal de entrada es de 50 mV a 25ºC sabiendo que I_B =10 nA a esa temperatura.

b) Calcule el error a 65ºC si la deriva térmica de la corriente de polarización es de 10 pA/ºC.

SOLUCIÓN:

El circuito equivalente del puente en funcionamiento es el presentado en la figura 3.10. El desequilibrio de R_S es la diferencia entre R_{S1} y R_{S2}, resistencias por las que circulan las corrientes de polarización del amplificador.

a) La tensión de error debida al desequilibrio máximo de R_S cuando el amplificador está a 25ºC será $v_{io} = I_{B\,25ºC}\,\Delta R_S = 10$ nA·500 Ω = 5 μV. El error relativo que esta tensión representa es 5 μV/50 mV, es decir, un 0,01%.

[4] Si se quiere disminuir el efecto de la corriente de desviación, se podrían disponer resistencias adicionales entre las patillas de entrada y masa pero eso contribuiría a reducir la tensión que entraría al amplificador al establecerse un divisor resistivo con la resistencia de salida de la fuente (no ha sido indicada por simplificar en la figura 3.9).

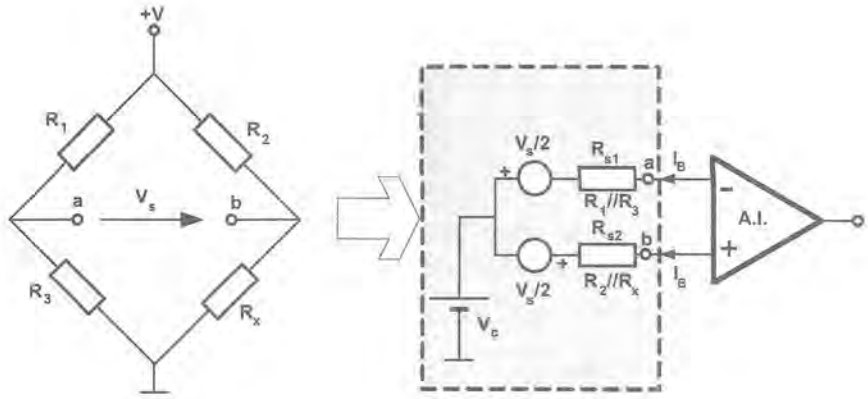


Figura 3.10

b) La corriente de polarización a 65°C será de 10 nA + 10 pA·(65–25)°C, es decir, 10,4 nA. Por tanto, la tensión de error debida al desequilibrio máximo de R_S cuando el amplificador está a 65°C será $v_{io} = I_{B\,65°C}\,\Delta R_S = 10{,}4\text{ nA}\cdot 500\ \Omega = 5{,}2\ \mu\text{V}$. Esto representa un error relativo de 0,0104%.

---

9. **Tiempo de establecimiento (*settling time*):** este parámetro —ya definido en el capítulo 1— caracteriza la respuesta temporal del amplificador y se proporciona como dato habida cuenta del uso de este tipo de dispositivos como etapa de entrada que lee la señal producida por el sensor. Se define en términos del tiempo necesario para que la salida se diferencie sólo en un 0,1%, o en un 0,01% del valor final. Es dependiente del ancho de banda, y por ende, de la magnitud de la ganancia. Se expresa para cada valor de ganancia y generalmente aumenta al incrementar ésta. Por ejemplo, para el amplificador PGA 204 de Burr-Brown se tiene que el tiempo de establecimiento es:

   ST (0,1%) = 22 μs para G = 1.
   = 100 μs para G = 100.
   ST (0,01%) = 23 μs para G = 1.
   = 140 μs para G = 100.

10. **Producto ganancia-ancho de banda (GBW):** corresponde a la misma idea que en el caso de los amplificadores operacionales y establece el límite de trabajo del amplificador de instrumentación. Los amplificadores de instrumentación no suelen poseer un GBW elevado ya que la mayoría de ellos están concebidos para funcionamiento en baja frecuencia o en continua, estando optimizado su diseño para este modo de trabajo.

11. **Razón de rechazo de modo común (CMRR):** nuevamente, es el mismo concepto que en el caso de los amplificadores operacionales e introduce el cambio que existe en la salida debido a la tensión de modo común (ambas entradas del amplificador soportan tensión de igual magnitud, signo y forma de variar con el tiempo). Se expresa, igual que para operacionales, como cociente entre la ganancia diferencial, A_d y la ganancia al modo común, A_c expresada en dB. Su comportamiento en frecuencia no es constante y desciende (a veces, fuertemente) al aumentar la frecuencia de la señal de modo común. La tensión que produce en la salida se puede obtener de la expresión:

$$v_{oc} = \frac{A_d v_c}{\log^{-1}\left[\dfrac{CMRR(dB)}{20}\right]} \tag{3.10}$$

EJERCICIO 3.6

*El amplificador de instrumentación del ejercicio 3.5 trabaja con una ganancia diferencial de 100 y posee una razón de rechazo al modo común para esta ganancia de 100 dB. Si la tensión respecto a masa del puente en los puntos **a** y **b** cuando $V_S = 0$ V es $V_c = 5$ V, ¿qué tensión existirá en la salida?*

SOLUCIÓN:

Haciendo uso de la expresión (3.10), tenemos que la tensión de modo común que se introduce en el sistema es de:

$$v_{oc} = \frac{A_d v_{ic}}{\log^{-1}\left[\frac{CMRR(dB)}{20}\right]} \Rightarrow v_{oc} = \frac{100 \cdot 5}{\log^{-1}\left[\frac{100}{20}\right]} = \frac{500}{10^5} = 5\ mV$$

lo que producirá el correspondiente error de salida.

Existen otros parámetros, tales como la no-linealidad de la ganancia, la tensión y corriente de ruido, el ancho de banda de potencia, el *slew-rate*, etc. que mantienen exactamente la misma interpretación que para un amplificador de propósito general. Otros parámetros como el margen de tensión de alimentación, la corriente de reposo, la razón de rechazo a los cambios en la alimentación, etc. tienen una interpretación muy directa, son también iguales a los de los amplificadores operacionales y no haremos referencia a ellos.

EJERCICIO 3.7

Un amplificador de instrumentación de ganancia programable por software cuyos parámetros se ofrecen en forma de tabla es utilizado como acondicionador analógico para un sensor. Si la señal de entrada posee un margen estático de: 0 ÷ 625 mV y la fuente de señal posee una resistencia interna $R_S = 500\ \Omega$:

a) Especifique la ganancia a seleccionar para lograr un margen estático de salida de 0 – 10 V.

b) Trabajando en el margen de temperatura de +25 a +85°C, calcule el error relativo de la salida en el peor de los casos debido a la no-idealidad de los parámetros ganancia y ruido del amplificador.

Características eléctricas del amplificador de instrumentación

PARÁMETROS	GANANCIA				
	1	2	4	8	16
Error de ganancia (%)	0,05	0,05	0,1	0,15	0,15
No linealidad a 25°C (%)	0,05	0,05	0,1	0,15	0,15
Deriva del error de ganancia (ppm/°C)	0,5	0,5	0,5	0,5	1
Tensión de ruido (RTI) @ 0,1 – 10 Hz (mV_{pp})	5	5	5	4	3
Corriente de ruido @ 0,1 – 10 Hz (nA_{pp})	60	60	60	60	60
Ancho de banda a 3 dB (MHz)	4	2	1,1	0,65	0,35
Ancho de banda de potencia (MHz)	0,1	0,1	0,1	0,35	0,35
Slew rate (V/ms)	6	6	6	24	24

NOTA: Se han presentado sólo las características eléctricas correspondientes a las ganancias de 1, 2, 4, 8 y 16 aunque el amplificador es de 11 pasos.

SOLUCIÓN:

a) A_d = 10 V/625 mV = 16. En el caso en que no coincidiera con alguna de las ganancias que el amplificador posee, se selecciona la inmediata inferior, con el problema de no poder lograr la magnitud de la tensión de salida requerida. El uso eficiente de un amplificador de ganancia programable estriba en el hecho de que los márgenes estáticos de las señales de entrada que serán conectadas al mismo hayan sido adecuados para conseguir un margen de salida común. Generalmente la salida del amplificador va conectada a la entrada de un sistema de conversión analógico-digital y este convertidor posee un margen estático de entrada que debe ser aprovechado al máximo.

b) Detallaremos en forma de tabla, el cálculo del error máximo que introduce la no-idealidad de cada parámetro del amplificador a 25°C y a 85°C.

TIPO DE ERROR	Cálculo del error (ppm)	Error a 25°C	Error a 85°C
Error de ganancia (G = 16): 0,15%	$0{,}15/100 \cdot 10^6$	1500	1500
Deriva del error de ganancia: 5 ppm/°C	$5 \cdot (85 - 25)$	–	300
No linealidad de la ganancia: 0,005% a 25°C 0,01% a 85°C	 $0{,}005/100 \cdot 10^6$ $0{,}01/100 \cdot 10^6$	 50	 100

Por tanto, la no-idealidad de la ganancia produce errores de 1550 ppm a 25°C y de 1900 ppm a 85°C lo que representan errores de 0,155% y 0,19%, respectivamente.

Por otra parte, el ruido del amplificador de bajas frecuencias está formado por dos componentes: la tensión de ruido y la tensión que produce la circulación de la corriente de ruido por la resistencia de la fuente R_S, es decir, $I_n \cdot R_S$ lo que implica, para este caso: $v_n = 30\ nA \cdot 500\ \Omega = 1{,}5\ \mu V$.

Como son aleatorios y no están correlacionados entre sí, el total de la tensión de ruido se obtendrá aplicando lo que estudiamos en el capítulo 2. Por ello, el ruido total más probable puede calcularse como:

$$v_n = \sqrt{1{,}5mV^2 + 1{,}5\mu V^2} \cong 1{,}5\ mV$$

lo que representa un error de 1,5/625 = 0,0024 o, lo que es lo mismo, un 0,24%.

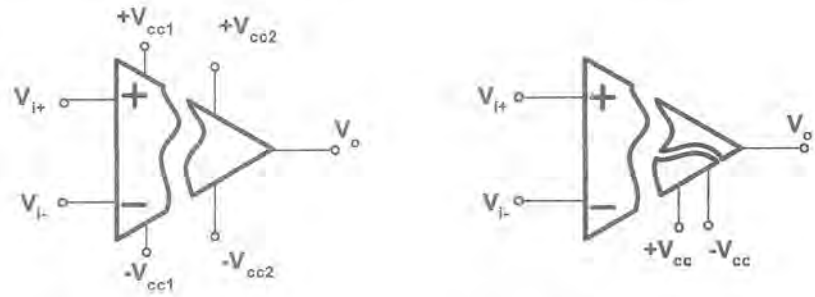
La magnitud de las tensiones de modo común para estos amplificadores de instrumentación puede superar a las fuentes de alimentación, siempre que se realicen conexiones tales que "eleven" el potencial del terminal de referencia respecto del de masa. Por ejemplo, el AD628 con fuentes de alimentación de ±15 V es capaz de trabajar con tensiones de modo común de hasta ±120 V, siempre que se eleve la tensión de su terminal V_{REF} a ese mismo valor. Este dato puede parecer un nivel muy alto, pero en muchas aplicaciones industriales la tensión de modo común puede alcanzar varios cientos de voltios y, además, ser muy variable en el tiempo, de tal forma que se superan las caracteristicas anteriores. Para tales aplicaciones se han desarrollado los amplificadores de instrumentación de aislamiento o aislados que veremos a continuación.

3.2 Amplificadores aislados o de aislamiento

La función esencial de estos amplificadores es lograr aislamiento galvánico entre sus circuitos de entrada y de salida mediante una **barrera de aislamiento**. Su empleo resulta imprescindible o resulta justificable en las siguientes situaciones:

- Cuando los niveles de tensión de modo común a la entrada del amplificador V_c son muy altos o se necesita amplificar señales provenientes de sensores en los que no exista aislamiento galvánico con tierra física en el punto de medición lo que puede originar tensiones de modo común muy elevadas entre esta tierra y la masa del circuito.
- Para evitar que los circuitos electrónicos del sistema de medida transfieran sus potenciales eléctricos a los sensores en contacto con el cuerpo humano (aplicaciones biomédicas).
- En otras aplicaciones en que se necesita un aislamiento entre los circuitos de entrada y los de salida del amplificador.

La figura 3.11 presenta el símbolo de este amplificador para sus versiones de 2 y 3 puertos. En los amplificadores de aislamiento de dos puertos, el dispositivo queda dividido galvánicamente en dos partes o "islas" aisladas entre sí, cada una de ellas alimentada independiente por fuentes de alimentación también aisladas entre sí (en los acoplados capacitivamente y los optoacoplados) o alimentados con una sola fuente externa, con su masa generalmente asociada a la del circuito de salida (en los acoplados inductivamente). En el amplificador de aislamiento de tres puertos, una sola fuente aislada de ambas partes del circuito (entrada y salida) alimenta mediante un pequeño convertidor CC/CC con transformador tanto a los circuitos de la entrada como a los de salida, garantizando así el aislamiento galvánico. Estos últimos pueden ser capaces de lograr aislamientos del orden de ±2500 V entre las masas de la entrada y la salida.



(a) Aislamiento de dos puertos

(b) Aislamiento de tres puertos

Figura 3.11. Símbolo del amplificador de aislamiento.

3.2.1 TIPOS DE AMPLIFICADORES AISLADOS

El acoplamiento entre los circuitos de entrada y salida para la señal a amplificar tiene como misión pasar la información de uno a otro lado sin que se precise la circulación de corriente eléctrica. Esto puede hacerse mediante aislamiento eléctrico, magnético o una combinación de ambos: electromagnético. Así, y de forma correspondiente, el paso de la información se puede realizar mediante técnicas capacitivas, inductivas u ópticas. En algunas ocasiones, el paso se hará "guardando la información" en una señal diferente (modulación) mientras que, en otros casos, será la propia señal de información (o una copia amplificada) la que excite directamente al dispositivo de transmisión.

La figura 3.12 presenta el diagrama simplificado de un amplificador de aislamiento de acoplamiento capacitivo. Este aislamiento emplea circuitos moduladores digitales, bien por ancho de pulso (PWM - *Pulse Width Modulation*) como en los modelos ISO103, ISO107 e ISO113 de Burr Brown, o por modulación de frecuencia (FM) como el ISO102 o el ISO106 también del mismo fabricante.

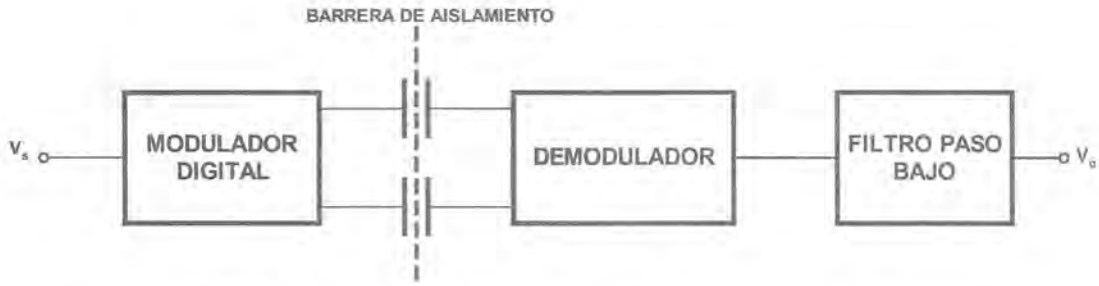


Figura 3.12. Diagrama simplificado de un amplificador aislado capacitivamente.

La señal de salida del oscilador interno es modulada mediante la señal de entrada y de esta forma es transferida (a través de los condensadores que conforman la barrera de aislamiento) al circuito de salida, donde es demodulada y filtrada para recuperar la información original.

La figura 3.13 representa el esquema simplificado del ISO102 en el que un oscilador controlado por tensión (VCO), cuya frecuencia para $v_i = 0$ es f_o, cambia dicha frecuencia en función de la magnitud y el signo de v_i, transfiriendo esta onda rectangular en forma diferencial a través de los condensadores de 3 pF que conforman la barrera de aislamiento. Esta señal es decodificada por el PLL (*Phase Locked Loop* o bucle de enganche de fase) que incorpora un VCO idéntico al de entrada, obligado a trabajar a la misma frecuencia y fase, con lo que la tensión de salida será muy parecida a la de entrada (idealmente análoga), pero aislada galvánicamente de ésta.

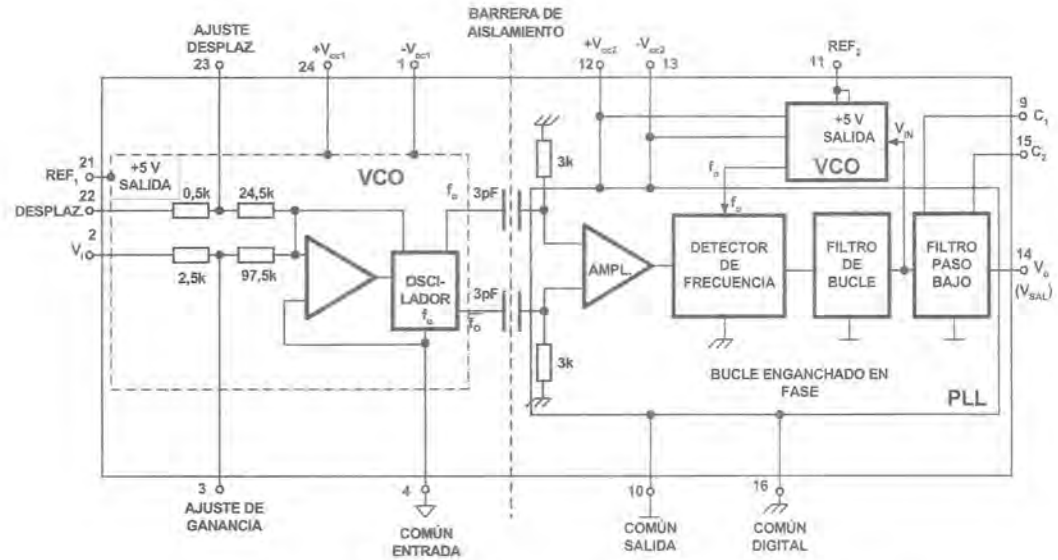


Figura 3.13. Esquema simplificado del amplificador aislado ISO102.

Los condensadores de aislamiento son muy especiales, formados por películas delgadas de wolframio depositadas en espiral sobre un sustrato cerámico de alta pureza. Esto posibilita obtener una rigidez dieléctrica superior a 10 kV y resistencias del orden de 10^{14} Ω.

Estos amplificadores permiten altas tensiones de aislamiento (entre 1500 y 2000 V), anchos de banda de pequeña señal bastante grandes (del orden de 70 u 80 kHz) y muy buena precisión de la ganancia a un coste relativamente bajo. Sin embargo, parte del rizado de la portadora aparece en la salida, produciendo errores a la salida y que son más acentuadas en los moduladores de tipo PWM. Además, aparece un ruido de banda ancha generado por el "temblor" (*jitter*) provocado por el proceso de modulación [3.4]. En el CD adjunto se proporciona las hojas de características del ISO102 y del ISO106, donde pueden apreciarse algunas aplicaciones típicas de este amplificador.

La figura 3.14 presenta el esquema simplificado de un amplificador aislado ópticamente. La señal de salida de la "isla aislada de entrada" en forma de corriente excita un diodo emisor de luz (LED) óptimamente acoplado a 2 fotodiodos. El diodo D_1 se utiliza para realimentar el amplificador de entrada y conseguir buena linealidad en el amplificador. El diodo D_2 recibe una señal idéntica a la que recibe D_1 por lo que la corriente relacionada con la señal de entrada en ambos lados de la barrera es igual y la función de transferencia del circuito depende en gran medida del acoplamiento óptico, con lo que garantiza buena linealidad, gran rapidez de respuesta y un ancho de banda moderadamente alto.

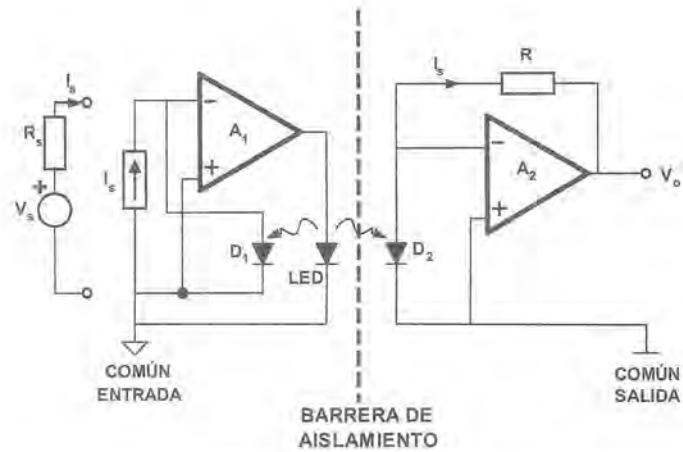


Figura 3.14. Esquema simplificado del amplificador aislado ópticamente.

La señal es amplificada en la etapa de salida, que suele ser un convertidor corriente-tensión (I/V) de resistencia R. Esta resistencia es externa en los casos en que la ganancia diferencial se controle en el circuito de la salida (como en el ISO100 de Burr-Brown) o es interna (generalmente 1 MΩ) y la ganancia diferencial se determina por resistencias $R_{G1,2}$ colocadas en el circuito de entrada (como en los BB 3650-3652). Estos amplificadores exhiben muy buena linealidad y estabilidad de la ganancia (0,02% y 0,005%/°C), un moderadamente alto aislamiento entrada-salida (±800 V) con un ancho de banda en pequeña señal y *slew-rate* similares a los de acoplamiento por condensador (60 kHz y 0,3 V/μs). Las hojas de datos del fabricante de estos amplificadores se ofrecen en el CD adjunto.

La técnica de acoplamiento óptico puede utilizarse para conseguir amplificadores aislados construidos con componentes discretos en los que el doble optoacoplador puede llegar a incorporar fibra óptica con lo que se consiguen ventajas añadidas en cuanto a tensión de aislamiento y distancia entre la etapa de entrada y la de salida. También es posible obtener anchos de banda muy elevados empleando componentes discretos de alta frecuencia llegando a obtener valores de GBW cercanos a 1 GHz [3.9].

La figura 3.15 presenta el esquema simplificado del amplificador de aislamiento mediante transformador AD202/204 de Analog Devices. El paso de la señal del circuito de entrada al de salida se realiza tras modular una portadora senoidal, en frecuencia con la señal, pasarla a través de un transformador de acoplamiento

y demodularla en el circuito de salida. Aquí el transformador constituye la barrera de aislamiento entre los circuitos de entrada y salida y es este elemento el que impone la principal limitación del ancho de banda del circuito (sólo 5 kHz con ganancia unitaria) aunque se consiguen tensiones de aislamiento de hasta ±2000 V con una buena linealidad (0,025%), mientras que la tensión de desviación (*offset voltage*) máxima referida a la entrada es de (±10 ± 10/G) mV[5].

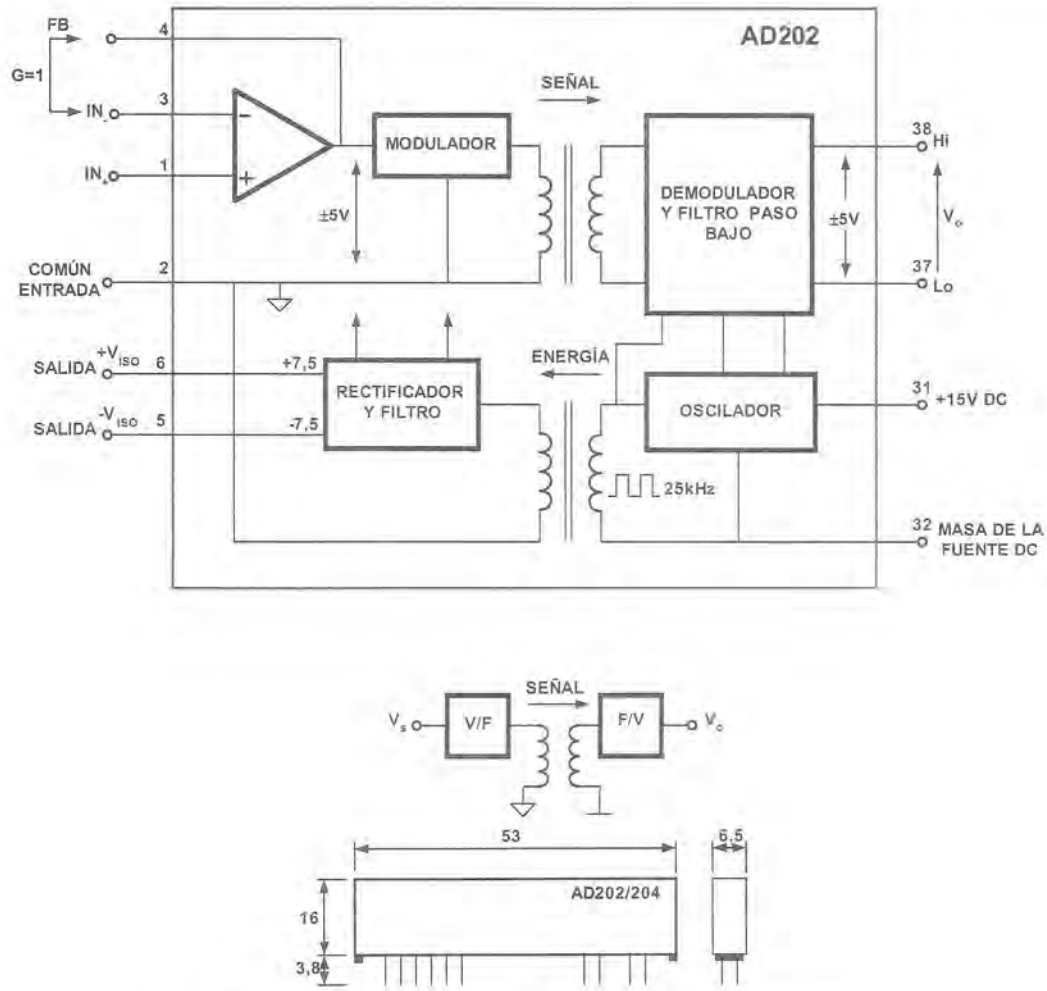


Figura 3.15. Amplificador aislado a transformador AD202.

Este circuito tiene una fuente de continua asociada a la etapa de salida que proporciona alimentación a los circuitos de entrada mediante un transformador adicional, por lo que el dispositivo es de dos puertos. La figura 3.15 presenta una vista simplificada con las dimensiones de este circuito en que se puede apreciar que no se trata de un dispositivo monolítico, sino que se trata de un bloque híbrido donde se combinan diversas tecnologías de fabricación en el mismo encapsulado. Obsérvese también la separación física que existe entre las patillas de la entrada y las de la salida y que debe ser lo suficientemente grande como para soportar la elevada tensión aplicada entre ellas sin que se produzca un arco eléctrico.

[5] En este circuito la ganancia de la etapa de salida es 1. Pero al referir el *offset* a la entrada de esa etapa (RTI) se divide por la ganancia de la etapa de entrada, G. Si G = 1, en el peor de los casos la tensión de *offset* será de ±20 mV.

Una gran parte de los amplificadores de aislamiento ofrecen en su circuito de entrada fuentes de tensión dobles referidas a la masa del circuito de entrada para alimentar puentes de Wheatstone con sensores o sensores directamente. En la figura 3.16 se muestra una aplicación con dos amplificadores aislados de este tipo que leen la señal de dos sensores incluidos en sendos puentes de Wheatstone —aislados entre sí y alimentados por el circuito amplificador desde la salida— y que ponen las señales en un circuito de medida.

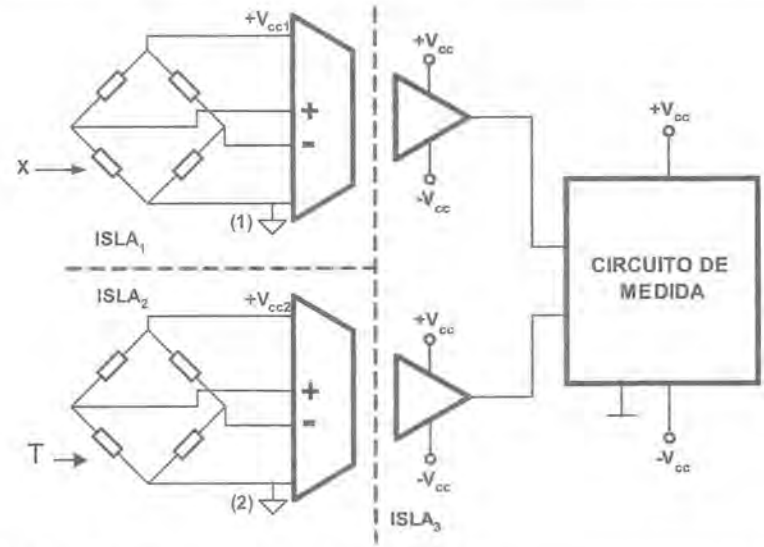


Figura 3.16. Aplicación de amplificadores aislados alimentando puentes de sensores.

3.2.2 PARÁMETROS CARACTERÍSTICOS DE LOS AMPLIFICADORES DE AISLAMIENTO

En este apartado nos centraremos en los parámetros más destacados de los amplificadores de aislamiento que no son comunes con los demás tipos de amplificadores que hemos estudiado hasta ahora.

La figura 3.17 presenta el diagrama de circuito de un amplificador de aislamiento. La tensión V_{ISO} es la que debe soportar el dispositivo entre la entrada y la salida (en la barrera de aislamiento), es decir, entre las masas de entrada y salida y la contribución como error a la salida de esta tensión se expresa mediante la **razón de rechazo del modo de aislamiento** IMRR (*Isolated Mode Rejection Ratio*), parámetro exclusivo de estos amplificadores y que viene a representar el efecto de la tensión que soporta el dispositivo entre la entrada y la salida sobre su funcionamiento.

El comportamiento de la barrera de aislamiento se representa mediante componentes parásitos que permiten la existencia de fugas entre la entrada y la salida: el modelo consiste en una impedancia, Z_{ISO} (está formada por una resistencia, R_{ISO} y una capacidad, C_{ISO}) por la que circula una pequeña corriente de fugas de la barrera I_L (idealmente debería ser cero). Pero los valores reales de estos parámetros permiten la circulación de corrientes que pueden llegar a ser de centenas de nanoamperios, aplicando la tensión de la red eléctrica como tensión de ensayo.

En las aplicaciones en que la tensión de modo común es muy elevada, la fuente de señal se conecta aislada de la masa de entrada y entonces, la tensión de modo común se refiere a la masa de salida, con lo que se aplica en su totalidad a la barrera de aislamiento. De esta manera, el amplificador es capaz de soportar tensiones de modo común del orden de miles de voltios [3.7].

Cuando la señal de entrada es una corriente i_s la tensión de salida del amplificador de la figura 3.17 puede expresarse en función de estos parámetros por la expresión:

$$v_o = i_s \cdot R_F \pm \frac{V_{ISO}}{IMRR} \tag{3.11}$$

y cuando la señal de entrada es una tensión v_s y tenemos un modo común aplicado v_c, la salida vendrá dada por:

$$v_o = \left(v_s \pm \frac{v_c}{CMRR} \right) \cdot \frac{R_F}{R_s} \pm V_{ISO} \cdot IMRR \cdot R_F \tag{3.12}$$

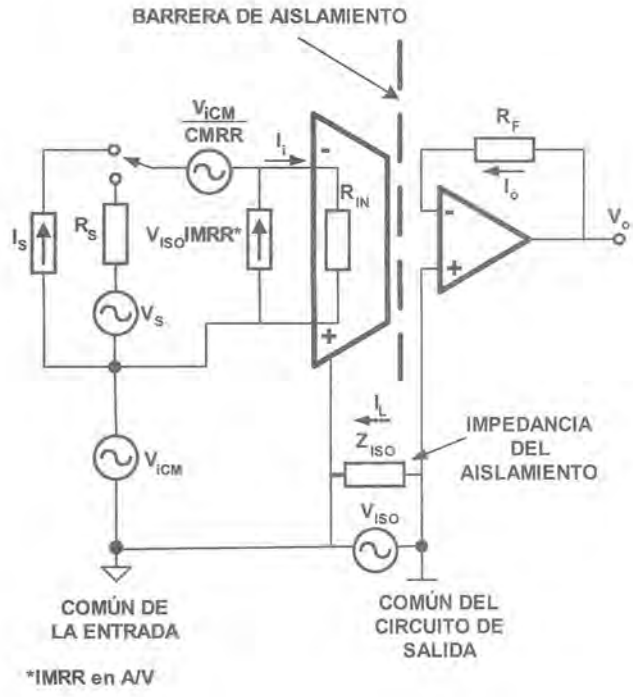


Figura 3.17. Diagrama circuital general de un amplificador de aislamiento.

En ambos casos, las expresiones (3.11) y (3.12) proporcionan una tensión de salida que representa la suma de la señal que se transmite y un error debido a la IMRR y/o CMRR.

La expresión (3.12) es válida cuando IMRR se exprese en términos de conductancia (A/V), mientras que si IMRR se expresara en V/V, entonces la expresión (3.12) se transforma en:

$$v_o = \left(v_s \pm \frac{v_c}{CMRR} \right) \cdot \frac{R_F}{R_s} \pm \frac{V_{ISO}}{IMRR} \tag{3.13}$$

Los valores de CMRR e IMRR pueden verse afectados por el desequilibrio entre las resistencias de salida, R_{s1} y R_{s2} que la fuente de tensión de entrada presenta sobre cada una de las entradas del amplificador según puede verse en la figura 3.18a.

En ambos casos, los máximos valores de CMRR e IMRR se consiguen cuando ambas resistencias son iguales mientras que CMRR e IMRR caen a medida que crece el desequilibrio entre ellas. Este dato se suele proporcionar en gráficos dentro de las hojas de características que el fabricante proporciona y pueden tener un aspecto como el de la figura 3.18b.

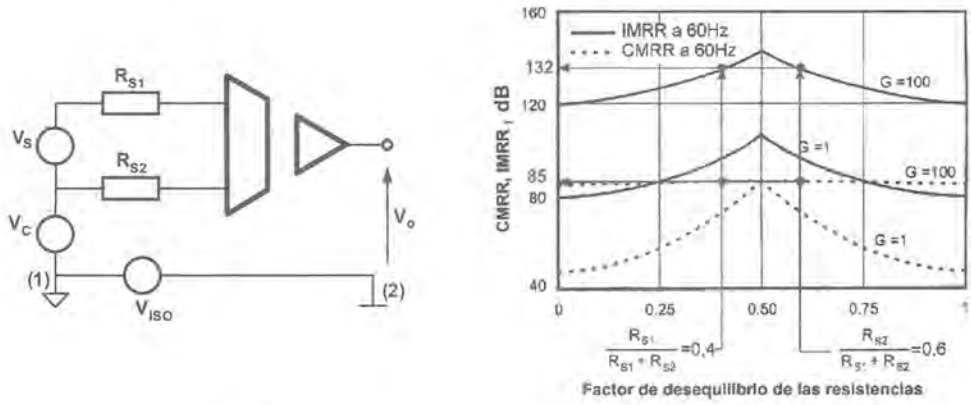


Figura 3.18. a) Resistencias en el circuito de entrada R_{S1}, R_{S2}, b) razones de rechazo CMRR, IMRR en función del desequilibrio de R_{S1} y R_{S2} para el amplificador optoaislado 3650.

EJERCICIO 3.8

Utilizando el amplificador optoaislado 3650 de Burr-Brown (sus características pueden comprobarse en el CD adjunto) con resistencias $R_F = 1\ M\Omega$.; $R_s = 10\ k\Omega$ con un desequilibrio máximo de 2 kΩ, calcule el valor de la tensión de salida cuando la tensión de entrada $v_s = 0$ V, es decir, el error absoluto que se produce por CMRR e IMRR. La tensión de modo común es de 60 V en continua y la tensión de aislamiento es de 440 V, también de continua.

NOTAS:

- *La expresión de la ganancia diferencial proporcionada por el fabricante es:*

$$A_d = \frac{10^6}{R_{s1} + R_{s2} + R_i} = \frac{10^6}{R_{s1} + R_{s2} + 25}$$

- *Considérese la figura 3.18b para el cálculo de los valores de CMRR e IMRR.*

SOLUCIÓN:

Por un lado, aplicando la ecuación del fabricante, la ganancia diferencial será (dado que $R_{s1} + R_{s2} = R_s$):

$$A_d = \frac{10^6}{4 \cdot 10^3 + 6 \cdot 10^3 + 25} = 99,75$$

Con el desequilibrio máximo de 2 kΩ en las resistencias de salida de la fuente, debemos obtener el cociente $R_{s1}/(R_{s1} + R_{s2})$ que es el que controla el valor de IMRR y CMRR (según la figura 3.18b): supongamos el caso de máximo desequilibrio en R_s que hace que $R_{s1} = 6$ kΩ y $R_{s2} = 4$ kΩ (o viceversa). El cociente anterior vale 0,6 (o 0,4).

Los valores de IMRR y CMRR para 0,6 (o 0,4) y una ganancia de 99,75 (consideramos 100), entrando en la gráfica de la figura 3.18b, son:

- IMRR = 132dB, lo que equivale aproximadamente a IMRR = 3980000
- CMRR = 85dB, lo que se traduce en CMRR = 17800.

Con estos valores, la tensión de salida —aplicando (3.13)— que se obtendrá con entrada nula será:

$$v_o \cong \left(0 \pm \frac{60}{17800}\right)\frac{1M}{10K} \pm \frac{440}{3980000} = \pm 0{,}337 \pm 0{,}000111\ V^{6}$$

Si la tensión de modo común se aplicara respecto a la masa de salida (al no referir a masa de entrada el circuito formado por la fuente de señal), entonces la tensión de modo común podría aumentar hasta 440V y obtener una salida de modo común más pequeña.

3.2.3 APLICACIONES DE LOS AMPLIFICADORES AISLADOS

Los amplificadores aislados son dispositivos costosos, sobre todo los que presentan tensiones de aislamiento muy altas y los de tres puertos con aislamiento por transformador. Además, la necesidad de acudir a tecnologías híbridas de fabricación hace que los precios no vayan a bajar demasiado en el futuro. Por todo esto, no son dispositivos que se vean muy a menudo en los sistemas electrónicos aunque sí que son bastante frecuentes en los sistemas de instrumentación. Veamos ahora cuáles son las principales aplicaciones en las que está perfectamente justificado el empleo de este tipo de componentes aun a pesar del incremento de coste a que dan lugar:

- Amplificación de señales de bajo nivel en presencia de tensiones de modo común muy elevadas, que requieren valores de la CMRR superiores a 120dB utilizando la técnica comentada al final del ejercicio 3.8.
- Interrupción de bucles de masa para hacer posible que los sensores queden aislados de los circuitos restantes del canal de medición. Esto también facilita la protección de instrumentos de medida de alta sensibilidad y la protección de los sensores y los organismos vivos en contacto con ellos (aplicaciones de biomedicina) ante las corrientes y tensiones de los circuitos del canal de medición.

Las limitaciones fundamentales de estos amplificadores radican en su pobre ancho de banda que —incluso en los más rápidos de acoplamiento óptico y capacitivo— sólo llega a unas decenas de kilohertzios a pequeña señal y aún menos en gran señal. La situación es aún peor en los amplificadores de aislamiento de acoplamiento magnético, por lo que sus aplicaciones se restringen a señales de espectro frecuencial reducido.

3.3 Amplificadores de transimpedancia y transconductancia

En el capítulo 2 estudiamos que el amplificador operacional se caracteriza por poseer una alta impedancia de entrada que admitía fuentes de tensión y una baja impedancia de salida que proporcionaba una excelente

[6] Realmente, el conciente R_f/R_s debería ser sustituido por la expresión de la ganancia del fabricante aunque, el error que se introduce es mínimo. El lector puede rehacer el cálculo más exacto y, seguramente, obtendrá valores muy parecidos a los aquí presentados.

fuente de tensión, es decir, se trataba de un buen amplificador de tensión. Sin embargo, existen otras realizaciones parecidas pero en las que o bien la entrada está prevista para admitir fuentes de corriente (impedancia de entrada muy baja) o bien la salida está diseñada como una fuente de corriente (impedancia de salida muy alta). En estos casos no puede hablarse de un amplificador de tensión, sino que estamos en presencia de amplificadores de transconductancia o de transimpendancia y su ganancia ya no será adimensional.

3.3.1 AMPLIFICADOR DE TRANSIMPEDANCIA

Como amplificador de propósito especial se ha desarrollado en forma de circuito integrado el denominado amplificador de transimpedancia o CFA (*Current Feedback Amplifier* o amplificador realimentado en corriente), cuya representación simbólica y circuito equivalente ideal se representan en la figura 3.19a mientras que su circuito real se muestra en la figura 3.19b en donde el parámetro Zm_{OL} se denomina **ganancia de transimpedancia** y tiene unidades de ohmios.

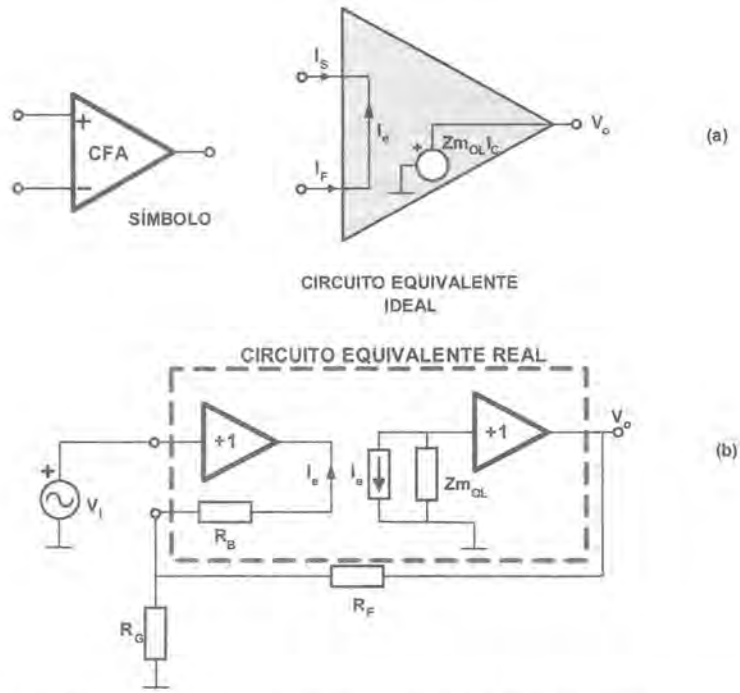


Figura 3.19. a) Símbolo y circuito equivalente ideal del CFA, b) circuito equivalente real.

Este amplificador difiere sustancialmente del operacional ya que su resistencia de entrada por la entrada inversora es muy pequeña (corresponde a la salida del *buffer* e idealmente sería cero) para posibilitar la libre variación de la corriente "inyectada" por la red de realimentación de corriente externa; mientras que la entrada no inversora presenta una moderadamente alta impedancia proporcionada por el *buffer* de ganancia +1.

El funcionamiento es el típico de una realimentación en corriente, como indica el propio nombre del dispositivo. El *buffer* fuerza a que la entrada inversora (–) del CFA "siga" las variaciones de la entrada no-inversora (+). Cuando la salida ha alcanzado el valor correspondiente a la entrada existente, la corriente de error i_e es muy pequeña (idealmente 0 si Zm_{OL} fuera infinitamente grande). Pero al producirse una variación en la entrada, se produce una significativa variación de i_e (posibilitada por la capacidad de drenar o entregar corriente del *buffer*) como resultado de la comparación entre la señal de entrada y la existente en la entrada

(–), proporcional a la tensión de salida. La muy alta ganancia en lazo abierto Zm_{OL} del CFA provoca una muy rápida variación de la tensión de salida hasta alcanzar el valor requerido en función de la ganancia (dependiente de las resistencias externas prácticamente) y de la magnitud de la señal de entrada.

El CFA se emplea para las mismas misiones que el amplificador operacional, entre las cuales, la más destacada es la de amplificar tensión. La configuración amplificadora más utilizada con el CFA es la de tensión no-inversora, presentada en la figura 3.20a.

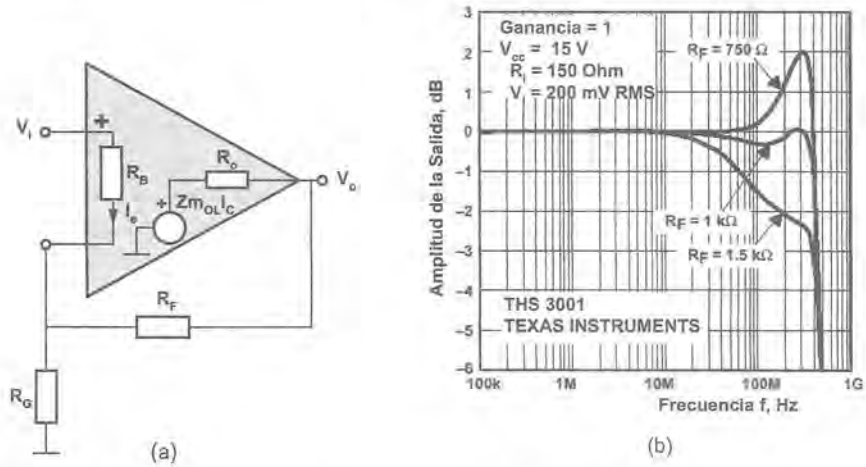


Figura 3.20. a) Configuración no-inversora, b) respuesta de frecuencia del CFA THS 3001.

La expresión de la ganancia de tensión para este amplificador, si R_B es muy pequeña (generalmente del orden de pocos ohmios), será:

$$\frac{v_o}{v_i} = \frac{Zm_{OL} \cdot \left(1 + \frac{R_F}{R_G}\right)}{R_F + Zm_{OL}} \tag{3.14}$$

Idealmente, con la ganancia de transconductancia infinita (o muy elevada), la expresión (3.14) para la ganancia de tensión del amplificador realimentado resulta idéntica a la del convencional amplificador operacional con realimentación de tensión y elevada impedancia de entrada, es decir:

$$\frac{v_o}{v_i} = 1 + \frac{R_F}{R_G} \tag{3.15}$$

El ancho de banda de este amplificador es muy grande pero depende, en gran medida, de R_F, disminuyendo al aumentar esta según se indica en la figura 3.20b. Sin embargo, modificando R_G se puede cambiar la ganancia de tensión sin alterar el ancho de banda —dentro de ciertos límites— lo que posibilita obtener valores muy elevados del producto ganancia-ancho de banda GBW al introducir una cierta independencia de la ganancia y del ancho de banda.

Sus elevadas ganancias, Zm_{OL} que pueden rondar el valor de 2 MΩ, su gran ancho de banda (del orden de cientos de megahertzios) y su *slew-rate* que puede llegar a ser de varios miles de voltios por microsegundo, los hace ideales para trabajar con señales de muy amplio espectro de frecuencias. Como ejemplo de uno de estos dispositivos, se adjunta en el CD las hojas de características del fabricante del CFA THS3001, producido por Texas Instruments.

Las aplicaciones fundamentales de este tipo de amplificadores se encuentran en los filtros analógicos pasobajo, integradores de señales de alta frecuencia y en amplificadores de señales de vídeo frecuencia donde se requiere un excelente comportamiento en frecuencia.

Sin embargo, su utilización con fuentes de señal de corriente no es adecuada, debido al incremento que sufre la figura de ruido con el incremento de la resistencia interna R_s. Por otro lado, los valores de las resistencias externas deben ser pequeños para poder mantener el ruido en la salida del amplificador en niveles bajos, debido a su gran ancho de banda.

3.3.2 AMPLIFICADOR DE TRANSCONDUCTANCIA

Un **amplificador de transconductancia,** también denominado fuente de corriente controlada por tensión (CSVC de las siglas *Current Source Voltage Controlled*), es el amplificador en el que la señal de entrada es una tensión y su salida una corriente. Existen amplificadores de propósito especial, similares a los operacionales, denominados de transconductancia (OTA, *Operational Transconductance Amplifier*) en los que se cumple esa condición. La figura 3.21 ofrece la representación y estructura interna de un OTA.

La ganancia de transconductancia del OTA es g_m y determina la corriente de salida I_o en función de la tensión diferencial de entrada:

$$I_o = g_m \cdot (v_{i+} - v_{i-}) \tag{3.16}$$

donde la g_m se obtiene como:

$$g_m = \frac{I_{ABC}}{2V_T}$$

donde I_{ABC} es la corriente de polarización, obtenida mediante una fuente de tensión V_1 y una resistencia R_B conectada en el terminal de polarización, es decir:

$$I_{ABC} = \frac{V_1 - V_{CC} - V_{D1}}{R_B}$$

y V_T es la tensión equivalente de temperatura del diodo.

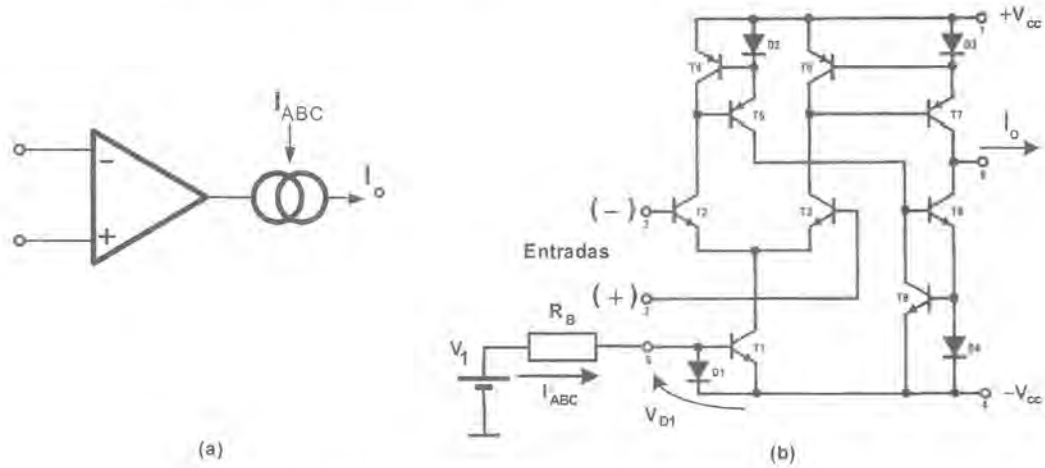


Figura 3.21. a) Símbolo del OTA, b) estructura interna.

La ganancia de transconductancia, g_m, puede alcanzar decenas de miles de microSiemens y se controla mediante la corriente de polarización I_{ABC} que alcanza algunos cientos de microamperios. En el CD adjunto ofrecemos las hojas de características del fabricante del LM13700 que incluye diodos en la entrada para incrementar la linealidad.

Si la magnitud de la corriente de salida deseada es mayor que i_o, algunos de estos circuitos integrados pueden incluir una etapa amplificadora de corriente (figura 3.22) formada por un transistor o un montaje Darlington, obteniéndose niveles de salida de varias decenas de miliamperios.

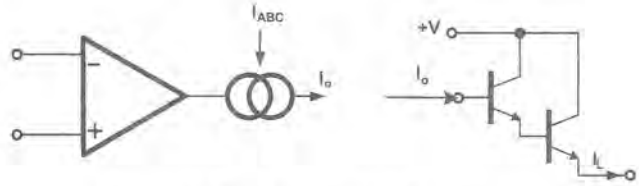
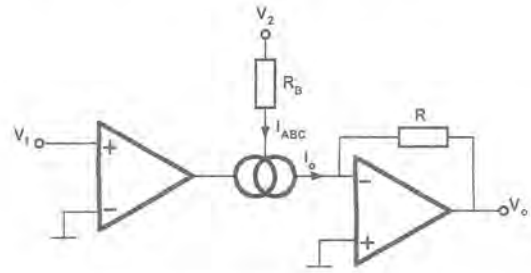
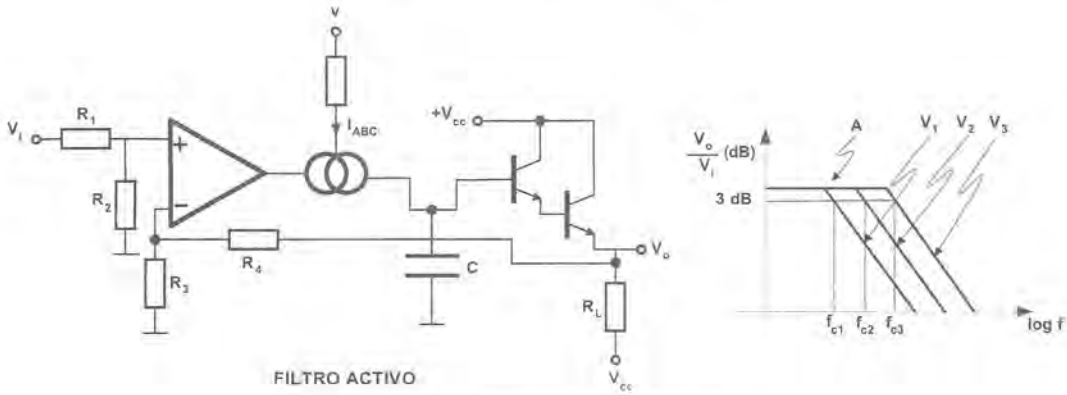


Figura 3.22. Amplificador de corriente.

Las aplicaciones fundamentales de los amplificadores de transconductancia se centran en multiplicadores analógicos, moduladores de amplitud y filtros activos. A modo de ejemplo, la figura 3.23 presenta un multiplicador analógico y un filtro paso bajo sobre la base de OTAs.



MULTIPLICADOR ANALÓGICO



FILTRO ACTIVO

Figura 3.23. Aplicaciones del OTA.

Para el circuito multiplicador analógico se tiene que:

$$V_o = -I_o \cdot R = -g_m \cdot V_1 \cdot R = -\frac{V_2 - V_{cc} - V_{DI}}{R_B} \cdot V_1 \cdot R = -\frac{R}{R_B} V_1 V_2 + \frac{V_{cc} - V_{DI}}{R_B} V_1$$

Obsérvese que la salida depende del producto V_1V_2, aunque para que el resultado sea exacto, I_{ABC} sólo deberá depender de V_2.

En el filtro activo paso bajo de frecuencia de corte controlada por la tensión V puede demostrarse que su frecuencia de corte es:

$$f_c = \frac{R_3 \cdot g_m}{(R_4 + R_3) \cdot 2\pi C} \tag{3.17}$$

Cuando v toma diferentes valores V_1, V_2, V_3 (donde $V_1 > V_2 > V_3$) se obtienen para el OTA ganancias de transconductancia $g_{m1} > g_{m2} > g_{m3}$ y, por consiguiente, se obtendrán frecuencias de corte $f_{c1} > f_{c2} > fc_3$ sin alterar la ganancia del filtro, lo que hace ideal este circuito para colocarlo a la salida de un amplificador de instrumentación de ganancia programable como acondicionador mediante multiplexación en el tiempo de varios sensores que requieran de anchos de banda diferentes.

3.4 Otros tipos de amplificadores

3.4.1 AMPLIFICADORES TROCEADORES

Una de las limitaciones de los amplificadores operacionales es la presencia en la salida junto a la señal de un conjunto de componentes de error directamente relacionadas con las tensiones de desviación, tensiones debidas a las corrientes de polarización, el ruido *flicker*, el modo común, así como la variación en el tiempo de estos errores debido a las derivas térmicas. Todos estos componentes pueden hacer que el error a la salida del amplificador bajo determinadas condiciones sea muy desfavorable. Si las señales a amplificar son muy débiles (corrientes del orden de picoamperios, tensiones de pocos microvoltios), los operacionales convencionales y los estudiados como amplificadores de instrumentación pueden no dar una solución válida.

Para reducir al mínimo estos errores se puede recurrir a la medición alternada de la señal y de los errores y a la elaboración de una salida "libre de errores". Este tipo de dispositivo se denomina como **amplificador operacional con autocero**.

La figura 3.24 presenta un esquema simplificado de la arquitectura interna de un amplificador operacional con autocero, o estabilizado mediante troceo. Hemos apuntado que se realiza la medición alternada de la señal y los "errores". La figura 3.24a presenta el estado del circuito en el modo de medición de los errores. El interruptor S_1 conecta las entradas de A_1, con lo que se cargará el condensador C_A con una tensión proporcional a los errores de las etapas A_1 y A_2, incluido el debido a CMRR no infinita de A_1. La duración de esta fase debe ser la mínima que permita que haya transcurrido el tiempo de establecimiento del sistema con realimentación negativa a través de A_3. Si los interruptores S_1 y S_2 pasan a la otra posición, el circuito se encontrará en el modo de medición de la señal. Ahora, a la señal de entrada se le restará el error que el circuito presentaba en la fase inmediatamente anterior, por lo que a C_B llegará la señal de entrada "limpia" de los errores del amplificador (solamente con los propios de la etapa final que representan muy poco sobre la señal amplificada). La señal a la salida de A_2 tendrá una apariencia "troceada", de ahí el nombre de **troceador** (*chopper*).

Durante el modo de auto-corrección o de medición de errores, la salida permanece en el mismo valor que tenía al finalizar la medición de la señal. Lo ideal es que la medición alternada de salida y errores se

pudiera hacer a una frecuencia f_M infinita para no influir en la forma de variar la señal con el tiempo, pero esto no es posible. Por ello, existirá distorsión armónica debida a la generación por parte del circuito de nuevas frecuencias en el espectro de la señal de salida.

Producto de la modulación que produce S_1 y la demodulación que produce S_2, en la figura 3.24b se presentan los espectros de amplitud de las señales en los puntos A y B. Es obvio que para obtener sólo el espectro de la señal en la salida v_o se necesitará filtrado, es decir, la eliminación de las componentes indeseables del espectro. C_B funciona, conjuntamente con la impedancia de salida de A_2 y con la impedancia de entrada de A_4, como un circuito de muestreo y retención (*sample & hold*). En la fase "auto-corrección" apenas pierde carga debido a la alta impedancia de entrada de A_4, manteniendo el valor de la señal a la salida, mientras que en la fase de "muestreo de la señal" su tensión "sigue" las variaciones de la señal con una muy pequeña constante de tiempo gracias a la muy baja impedancia de salida de A_2. Esto sirve como filtrado siempre que se seleccione bien el valor de C_B.

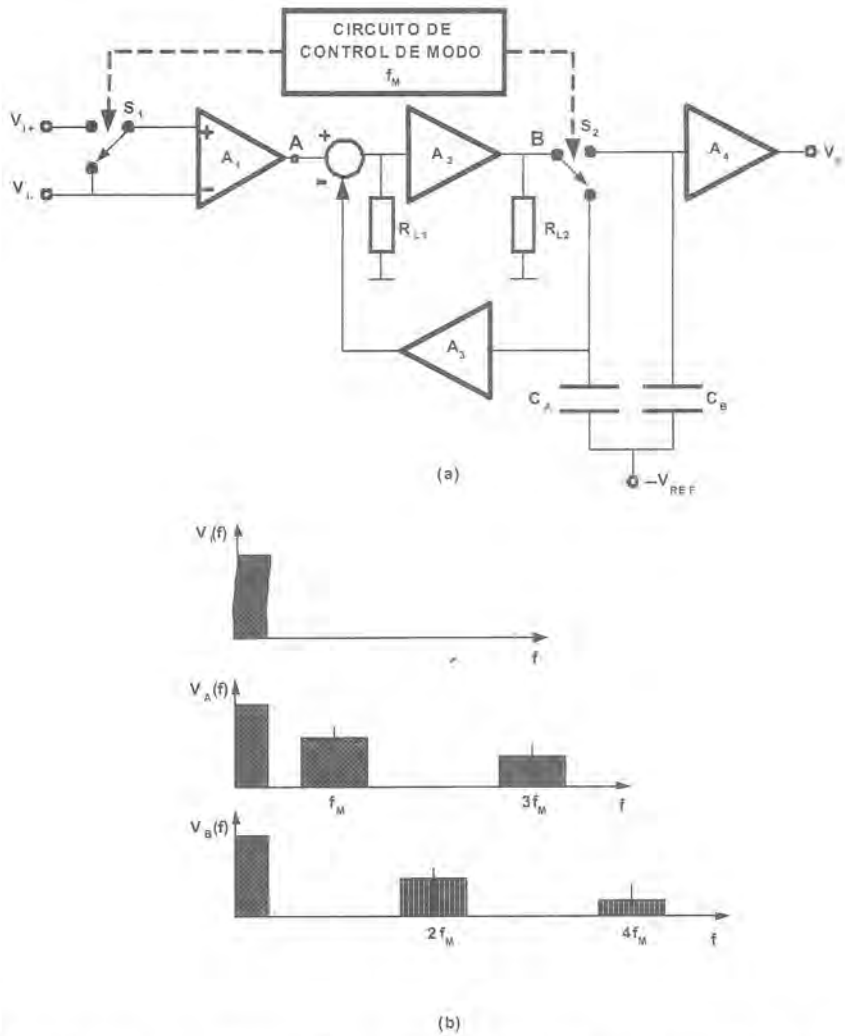


Figura 3.24. a) Esquema simplificado del amplificador de autocero, b) espectros de amplitud de la señal en A y B.

A modo de ejemplo, el TLC2652 fabricado por Texas Instruments, cuya hoja de características se ofrece en el CD adjunto, es un amplificador de precisión estabilizado mediante *chopper* que utiliza el principio expuesto. Posee un reloj interno de sólo 450 Hz, y una entrada para un reloj externo. En la medida que la frecuencia del reloj sea mayor, crecen las tensiones y corrientes de desviación, así como la corriente de polarización de entrada, empeorando drásticamente sus características. El producto ganancia-ancho de banda ronda los 2 MHz para $f_M = 10$ kHz.

Existen otros circuitos con reloj interno en el orden de 10 ó 20 kHz, posibilitando así el tratamiento de señales de amplio espectro de frecuencias. Estos circuitos son de gran aplicación en sensores que ofrecen muy pequeña señal como galgas extensométricas, termopares, sensores químicos, etc.

3.4.2 AMPLIFICADORES LOGARÍTMICOS

Un **amplificador logarítmico** es un amplificador de propósito especial cuya función de transferencia estática responde a la expresión:

$$y = \log x$$

donde **x** es la señal de entrada e **y**, la salida. La gráfica de esta función de transferencia para un amplificador logarítmico de entrada y salida de tensión tiene la forma presentada en la figura 3.25, en la que el eje de la señal de entrada es logarítmico con lo que se consigue una gráfica lineal para el amplificador ideal.

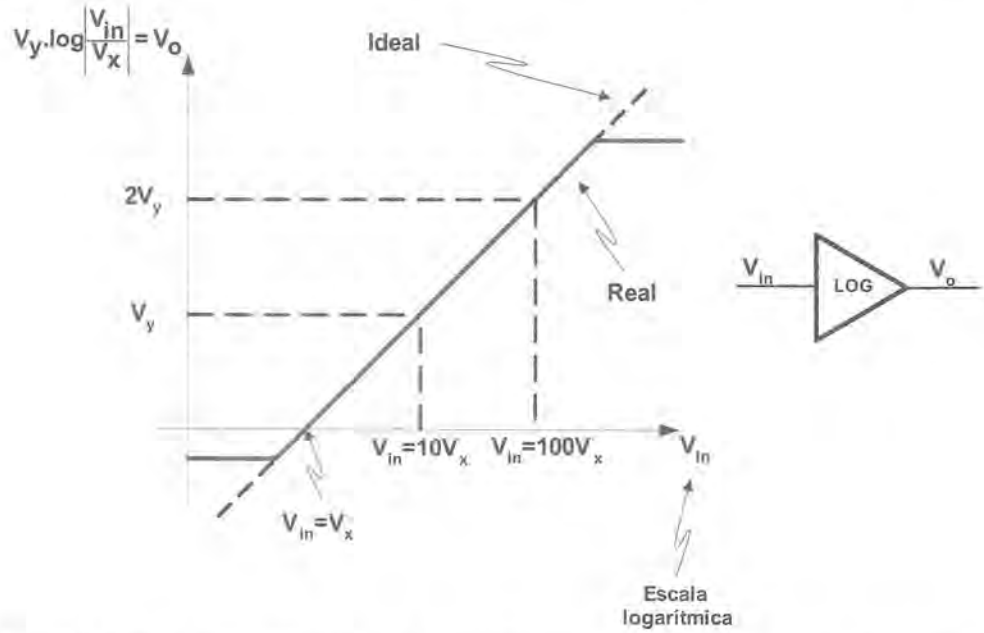


Figura 3.25. Símbolo y función transferencial ideal y real de un amplificador logarítmico.

En el amplificador real, por limitaciones conocidas, la salida no puede alcanzar altos valores, presentando un margen estático de salida limitado. En cambio el amplificador logarítmico es capaz de "comprimir" la información metiendo un mayor recorrido de datos en el margen de salida del operacional.

La tensión de entrada que hace nula la salida se denomina tensión de corte en cero, V_x, mientras que la tensión de salida V_y para $v_i = 10V_x$ se denomina tensión de pendiente (*slope voltage*) o factor de voltios por

década, y se expresa en V/dB. El comportamiento de los amplificadores logarítmicos se modela, pues, con dos referencias: V_x y V_y, con lo que la tensión de salida, v_o en función de la entrada, v_i será:

$$v_o = V_y \cdot \log_{10} \left| \frac{v_i}{V_x} \right| \tag{3.18}$$

Estos amplificadores se fabrican en forma monolítica en tres arquitecturas básicas: amplificador logarítmico con diodo, amplificador logaritmico de vídeo y el amplificador logarítmico de detección sucesiva.

La figura 3.26 representa el esquema básico del amplificador logarítmico con diodo en el que si consideramos que el operacional es ideal, la corriente que circula por el diodo i es la misma que la de entrada i_i; entonces, la tensión de salida v_o para corrientes muy superiores a I_o (siendo I_o la corriente de saturación inversa del diodo), vendrá dada por la expresión:

$$v_o = \frac{kT}{q} \ln \frac{i_i}{I_o} \cong 0,06 \log \frac{v_i}{R \cdot I_o} \tag{3.19}$$

Aun cuando en el circuito de la figura el diodo D sea implementado con un transistor, el margen dinámico de entrada para este circuito es pequeño (unos 80 dB) [3.1]; tanto la pendiente como el corte con cero de la función de transferencia son muy dependientes de la temperatura y sólo es capaz de trabajar con señales de entrada unipolares. Si consideramos una cascada de N etapas de amplificadores lineales similares limitados[7], cada uno de ganancia A (en dB) y con un comportamiento en gran señal bien definido, la ganancia total de la cascada será NA (en dB) para muy pequeña señal.

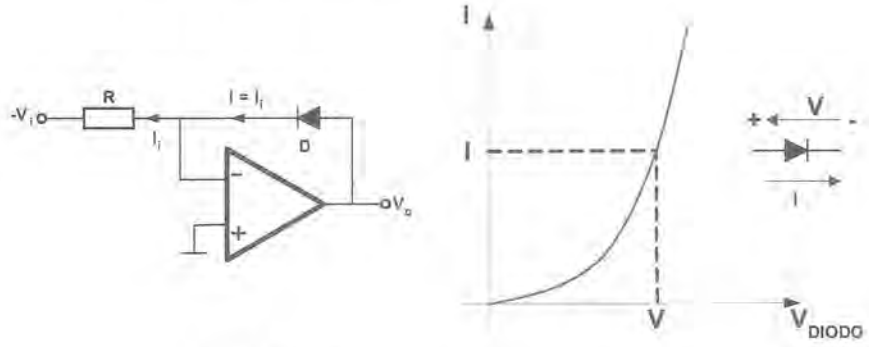


Figura 3.26. Amplificador logarítmico a diodo.

Si, como se muestra en la figura 3.27a, la salida de cada etapa se lleva a la entrada de un sumador se obtendrá la función de transferencia estática de la figura 3.27b.

Cuando la señal de entrada es lo suficientemente pequeña como para no llevar al límite a la última etapa, la salida (resultado de la suma) será muy parecida a la salida de la última etapa, e igual a NA (en dB). Cuando la señal de entrada aumenta y la etapa final llega a su límite, a partir de ahí no realiza contribución alguna a la salida por lo que la ganancia incremental de la cascada cae a (N – 1)A (en dB).Si la entrada continúa creciendo, la penúltima etapa llegará a su límite y a partir de este la ganancia de la cascada caerá a (N – 2)A (en dB). De esta forma, cuando la primera etapa llegue a su límite, la ganancia de la cascada será idealmente de 0 dB, y la gráfica salida-entrada de la cascada es un conjunto de líneas rectas que pueden constituir una buena aproximación a una curva logarítmica, donde la selección del valor de A es determinante en la linealidad. Si A es muy

[7] Un amplificador de tensión lineal limitado exhibe idealmente una característica de ganancia A desde señal nula hasta un cierto valor de ésta, $\pm V_{i\,MÁX}$, a partir del cual la salida permanece constante en un nivel.

grande, la cascada tendrá menos elementos, pero la aproximación será mala. Si A es pequeña (buena aproximación) entonces se requieren muchas etapas para conseguir un buen margen dinámico. Por lo general, en los circuitos integrados que se basan en este principio, se eligen ganancias entre 10 y 12 dB[3.1].

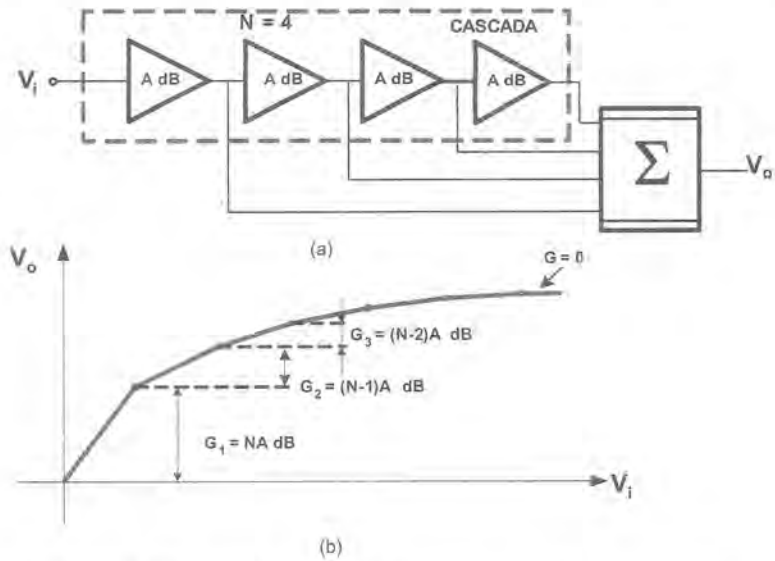


Figura 3.27. Amplificador logarítmico para señales unipolares.

La implementación de esta cascada con un buen funcionamiento a altas frecuencias resulta muy difícil, debido al retraso variable con la amplitud de la señal de entrada que el sistema introduce: para pequeñas señales el atraso será Nδt, mientras que para gran señal puede que sólo sea δt, siendo δt el retardo que aporta cada etapa de la cascada.

Una solución a este problema la aporta la arquitectura mostrada en la figura 3.28, en la que las etapas de la cascada en lugar de limitar la salida a partir de un cierto valor de la entrada, pasan a tener ganancia unitaria, con lo que resulta innecesario sumar las salidas de las etapas.

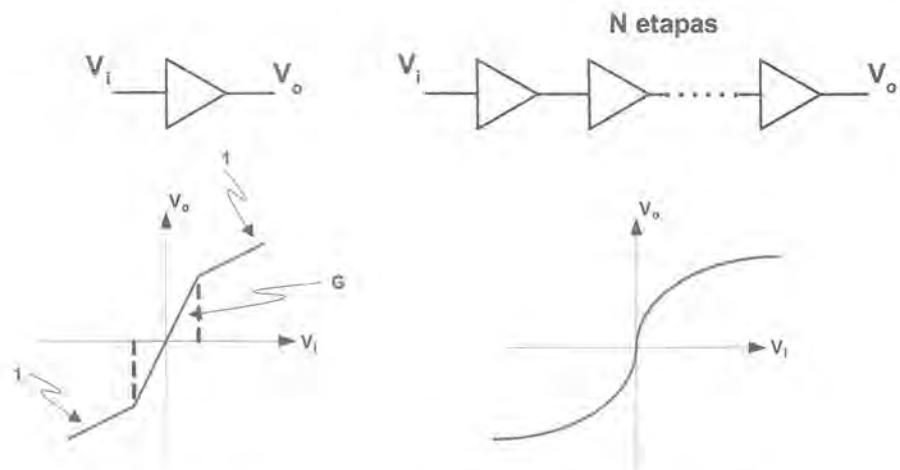


Figura 3.28. Cascada de N etapas con G = 1 para gran señal.

Ambas arquitecturas son utilizadas como amplificador logarítmico de vídeo, aunque la técnica más utilizada para altas frecuencias es la de detección sucesiva.

El amplificador logarítmico de detección sucesiva consiste de una cascada como la descrita anteriormente en la que en lugar de sumar las salidas de cada elemento directamente, estas son aplicadas a detectores cuyas salidas se suman.

La figura 3.29 ilustra el esquema de principio de este amplificador con 4 etapas de amplificadores lineales limitados.

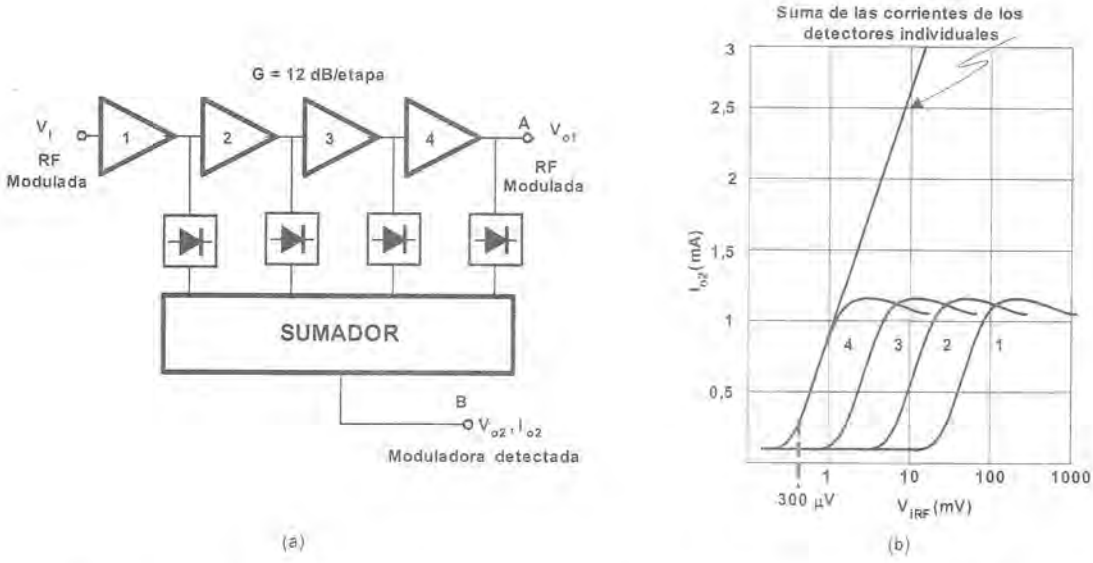


Figura 3.29. a) Amplificador logarítmico de detección sucesiva de 4 etapas de 12 dB, b) característica transferencial de la salida de corriente I_{o2}.

Si los detectores poseen salida de corriente, el proceso de suma se simplifica con la simple unión de sus salidas, que pueden ser rectificadores de media onda o de onda completa, mas no simples diodos. Los amplificadores que utilizan esta técnica ofrecen dos salidas: la limitada (A en la figura 3.29a) y la logarítmica (B en la misma figura). Obsérvese que en la salida del amplificador, en este caso de corriente, se obtiene la suma de las corrientes que aporta cada detector. De esta forma, la característica transferencial estática es una línea recta para entradas del amplificador desde 300 µV hasta 100 mV (figura 3.29b) [3.1], lo que representa un margen dinámico de entrada de alrededor de 48 dB. Los amplificadores de cada etapa poseen un gran ancho de banda para su uso fundamental en la detección o demodulación de portadoras de radiofrecuencia (RF) que pueden ser señales de cientos de megahertzios con información de amplio ancho de banda (como en el caso de vídeo que oscila entre 10 y 20 MHz).

Los parámetros fundamentales que caracterizan a un amplificador logarítmico pueden resumirse en los siguientes, algunos de los cuales tienen definiciones similares a las del resto de los amplificadores:

1. **Pendiente (*Slope*):** gradiente de la función de transferencia del amplificador, en V/dB o mA/dB.

2. **Tensión de corte en cero (*Voffset*):** valor de la señal de entrada para la cual la salida es cero, en dBV[8].
3. **Margen dinámico (*Dynamic range*):** cociente entre el valor máximo y el mínimo de la señal de entrada para el cual el amplificador ofrece una respuesta logarítmica, valor que se proporciona en dB. Por ejemplo, si la mínima y la máxima tensión que proporcionan respuesta logarítmica del amplificador son, respectivamente, ±40 μV y ±2,2 V o, lo que es lo mismo, el amplificador funciona entre –91 y +4 dBV, el margen dinámico es de 95 dB.
4. **Ancho de banda:** margen de frecuencia de la señal de entrada para el cual el amplificador funciona correctamente.
5. **Linealidad logarítmica (*Logaritmic lineality*):** mayor desviación de la característica de transferencia real respecto a una línea recta (en un diagrama logarítmico); este valor también se proporciona en dB.

A modo de ejemplo, el lector puede consultar los datos del amplificador logarítmico AD8310 fabricado por Analog Devices, cuyas hojas de características se encuentra en el CD adjunto.

Las aplicaciones de estos amplificadores en instrumentación son muy diversas: multímetros de valor real en decibelios para señales de alterna de alta frecuencia, analizadores de redes y de espectros de señales, procesadores de señales de radar y sonar de bajo coste, indicadores de nivel de potencia de señales radiadas por antenas de RF y la conversión de un nivel de señal a su expresión en decibelios.

3.5 Algunos criterios de selección de amplificadores operacionales para instrumentación

La selección de un operacional para un determinado circuito no es una tarea fácil. Un dispositivo que se ajuste perfectamente a las exigencias técnicas de todas las aplicaciones no existe en la práctica (si así fuese, cualquier problema sería muy sencillo de resolver y sobraría este capítulo y el anterior). Incluso en la mayoría de las aplicaciones no es conveniente que el amplificador sea totalmente ideal; por ejemplo, un amplificador de señales provenientes de sensores de temperatura no debería tener mucho ancho de banda ya que la señal es de continua y la reducción del ancho de banda, mejora la relación S/N, En este caso —como en otros similares— no es imprescindible la "idealidad".

Las diferentes tecnologías de fabricación de operacionales monolíticos tratan de optimizar algunos parámetros, pero resulta imposible "idealizarlos" todos con una misma tecnología, producto de las limitaciones inherentes a las particularidades físicas de la fabricación de transistores y resistencias.

Las tres tecnologías básicas de fabricación de los operacionales son: bipolar, BiFET y CMOS. En la tecnología bipolar todos los transistores del amplificador son npn o pnp (incluso los del circuito de entrada) por lo que el operacional tendrá bajos niveles de tensión de desviación y de ruido en la entrada pero un alto nivel de corriente de ruido; a su vez el operacional se caracteriza por una elevada ganancia en bucle abierto, A_d y *slew-rate* [3.5].

En contraste con esto, los operacionales diseñados con tecnología BiFET posee transistores de efecto de campo (FET) en el circuito de entrada y bipolares en el resto del circuito. Esto hace que presenten mejores

[8] En algunos casos, las tensiones se expresan en decibelios, agregando una "V" al símbolo. Por ejemplo: –88 dBV es equivalente aproximadamente a 40 μV (20log 0,00004 ≈ –88 dBV).

características de alta resistencia de entrada que los bipolares y *slew-rate*, pero son los que presentan niveles de tensión de desviación y de derivas térmicas más altos.

Hasta ahora, podemos apuntar que estas tecnologías dan como resultado operacionales de un gran ancho de banda y con una gran capacidad de corriente en la salida. Desde el punto de vista de la entrada, si los bipolares ofrecen la mejor solución para trabajar con fuentes de señal de bajas resistencias, para fuentes de alta resistencia interna es aconsejable el uso de un amplificador BiFET.

La tecnología CMOS para la construcción de operacionales es bien conocida por diseñar circuitos de muy bajo consumo y con la capacidad de trabajar con una sola fuente de alimentación, lo que los hace ideales para aplicaciones portátiles o de alimentación con baterías. También sabemos que los transistores CMOS de la entrada poseen cuasi-infinita resistencia de entrada (casi cero corriente de polarización), lo que los hace ideales en el sentido de la resistencia de entrada. Pero estos dispositivos requieren celdas de protección electrostática (ESD) en serie con las puertas aisladas de entrada que degradan estas características de impedancia. Poseen tensiones de desviación de valores intermedios entre los de las tecnologías anteriores, corrientes y tensiones de ruido muy pequeñas y muy altas derivas térmicas.

Con el desarrollo de la ciencia y la técnica han ido surgiendo nuevas tecnologías, como la DIFET (FET dieléctricamente aislado) que reduce casi a cero la corriente de polarización. Con esta tecnología se producen operacionales de precisión de bajo ruido, con características superiores a los de tecnología BiFET en casi todos los sentidos.

A modo de resumen, la figura 3.30 presenta la caracterización de los operacionales en función de sus tecnologías básicas de fabricación.

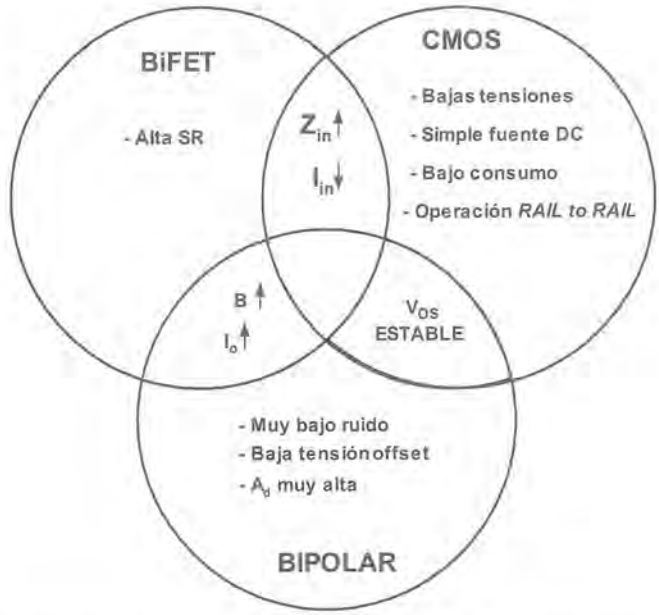


Figura 3.30. Características de las tecnologías de fabricación de amplificadores operacionales.

Las zonas comunes de los círculos son características en buena medida parecidas en orden de magnitud entre dos tecnologías (no existen características comunes en todas) que sugerimos sea por donde primero ubiquemos las principales exigencias de nuestra aplicación y luego tomemos la decisión en función de las restantes exigencias de menor peso.

Para conformar las exigencias de nuestra aplicación debemos considerar fundamentalmente los siguientes aspectos:

— Características de la fuente de señal y de la carga.
— Respuestas de amplitud, frecuencia y fase necesarias.
— Nivel de corriente de reposo (sin señal) y nivel de las tensiones de alimentación.
— Margen de temperatura de trabajo.

En el caso de los circuitos amplificadores hay que diferenciar dos tipos:

- **Amplificadores para señales referidas a masa.** La selección del operacional más adecuado se realizará en función del papel que tenga en el sistema de medida. Las etapas de entrada generalmente trabajan con señales muy pequeñas provenientes de sensores. Además, esta etapa tiene que generar muy poco ruido a su salida, tener tensiones y corrientes de desviación muy pequeñas y muy alta estabilidad térmica; no resulta importante que posea una gran ganancia. Es en esta etapa donde, en buena medida, se decide la calidad del sistema de medida. En función de la magnitud de la resistencia interna de la fuente de señal, podemos seleccionar entonces para ella un operacional BiFET o bipolar. Si no se encuentra una solución debido a que la señal es muy pequeña, entonces hay que recurrir a un operacional con autocero o estabilizado mediante *chopper*. Para etapas de salida o intermedias los bipolares o BiFET pueden ofrecer buenos resultados.

- **Amplificadores para señales diferenciales con la presencia de modo común.** En este caso, la etapa de entrada tiene que ser un amplificador de instrumentación que ofrezca una alta razón de rechazo al modo común (CMRR) y genere el menor ruido posible. Estos parámetros dependen básicamente de la tecnología de fabricación de su etapa de entrada.

  La configuración a seleccionar (diferencial, de dos operacionales o de tres operacionales) dependerá básicamente de los niveles máximos de modo común y del desequilibrio de la resistencia de la fuente. A partir de aquí se toman los mismos criterios que para las etapas intermedias y/o de salida (si son necesarias) que en el caso anterior. En los casos de necesidad de aislamiento y/o muy elevadas tensiones de modo común las opciones se circunscriben a los amplificadores aislados.

Por último, los fabricantes "clasifican" sus productos en función de los parámetros optimizados en el circuito, siendo la más generalizada la siguiente:

– Operacionales de baja corriente de entrada (menores de 5 pA).
– Operacionales de alta corriente de salida (mayores de 200 mA).
– Operacionales de bajo ruido (menores de 1 $nVHz^{1/2}$).
– Operacionales de muy bajo consumo (menos de 500 μA en reposo).
– Operacionales de bajas derivas térmicas (menores de 5 nV/°C y 0,01 pA/°C).
– Operacionales de gran ancho de banda (mayores de 100 MHz).

Esta clasificación puede darnos un primer resultado a la búsqueda del mejor operacional, aunque como hemos comentado hay que considerar otros factores.

3.6 Otras consideraciones en el diseño de circuitos con amplificadores operacionales

Cuando se diseñan circuitos con amplificadores operacionales hay que tener en cuenta algunas consideraciones adicionales, además de sus parámetros reales, que hacen que su funcionamiento se aparte del ideal. En

otras ocasiones, el funcionamiento de alguno de los circuitos de la tabla 2.1 se convierte en imposible debido a esas no-idealidades y debe ser modificada su topología para conseguir que su comportamiento se aproxime al previsto.

Vamos a tener en cuenta dos tipos de consideraciones:

Las resistencias que hemos considerado en todos los cálculos como de un valor exacto, nunca lo serán puesto que su proceso de fabricación impone tolerancias mayores o menores que introducen incertidumbre en cuanto a su valor. Esta incertidumbre se acota mediante un parámetro de tolerancia que indica cuánto puede separarse su valor del valor nominal de la resistencia. Así, tenemos resistencias del 5%, 2%, 1% o 0,1%.

Las resistencias también van a introducir derivas térmicas y su valor se modificará con la temperatura bien hacia valores más altos en el caso de resistencias de película metálica, bien hacia valores menores en el caso de resistencias de carbón. En general, se proporciona esa variación en valores por unidad (p.u.) por unidad de temperatura. Por ejemplo, podemos encontrarnos con una resistencia R de +15ppm/°C que indica que cada grado centígrado que varíe su temperatura, subirá su valor en 0,000015·R.

Los circuitos que incorporen condensadores se encontrarán con problemas del mismo tipo pero más graves puesto que las variaciones de valores de los condensadores y las derivas térmicas suelen ser aún mayores que en las resistencias.

Algunas de las topologías de la tabla 2.1 no pueden funcionar correctamente porque la presencia de no-idealidades en los amplificadores operacionales introducen cambios en su comportamiento que los hacen inviables como en el caso de los circuitos integradores.

Estos aspectos serán tratados a continuación.

3.6.1 CIRCUITOS BASADOS EN OPERACIONALES CON RESISTENCIAS REALES

En los circuitos en que las resistencias determinan la ganancia y no tienen que cumplir ninguna condición más, el efecto de los cambios en las resistencias respecto de su valor nominal sólo afecta al valor de la ganancia que resultará ser mayor o menor que la deseada; tal es el caso de los circuitos amplificadores inversor y no inversor y de los convertidores corriente-tensión y tensión-corriente de la tabla 2.1.

Sin embargo, hay otros circuitos de la tabla 2.1 tales como el amplificador diferencial o el rectificador de precisión en los que, para obtener una determinada función entrada-salida, se debe cumplir una determinada relación entre algunas resistencias. En tales casos, además del problema del cambio en la ganancia, aparece una cierta desviación del comportamiento previsto.

Consideremos, el caso del amplificador diferencial en el que dos resistencias deben ser iguales a R_1 y las otras dos, iguales a R_2 para conseguir que la expresión sea la que se indica en la tabla 2.1.

Supongamos ahora que no se da la condición de estricta igualdad en los valores y que, por tanto, en cada pareja, una de las resistencias tendrá más valor que la otra. En la figura 3.31 se muestra el amplificador en donde se han situado resistencias diferentes en cada rama.

La tensión de salida que se obtiene operando un poco resulta ser:

$$v_o = \frac{R_2 + \Delta R_2}{R_1 + \Delta R_1}(v_2 - v_1) + \frac{R_2 \Delta R_1 + R_1 \Delta R_2}{(R_1 + R_2)(R_1 + \Delta R_1)} v_2 \qquad (3.20)$$

lo que significa que se produce el previsible cambio de la ganancia diferencial pero —lo que es mucho peor— se introduce un término de modo común dependiente del valor de v_2 que puede llegar a ser muy importante e invalidar el trabajo del circuito. Los casos en que eso ocurrirá son:

1) Cuando se tenga un valor muy elevado de la tensión de modo común ya que una porción de ella aparecerá en la salida, mayor cuanto más discrepen entre sí las resistencias R_1 y R_2.

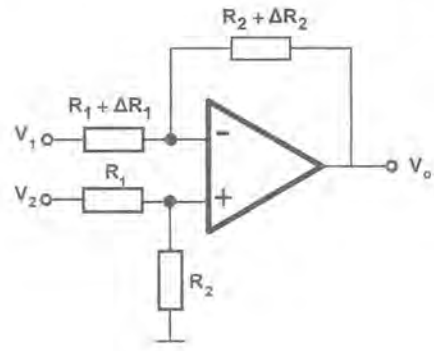


Figura 3.31. Amplificador diferencial.

2) Cuando se utilice el amplificador para eliminar interferencias de modo común. En estos casos, como en el ejercicio 2.10, una parte de esa interferencia aparece en la salida por causa de la CMRR y, otra parte, por causa de las discrepancias entre resistencias que deberían ser iguales.

La mejor solución para reducir el problema es hacer que las discrepancias ΔR_1 y ΔR_2 sean lo más pequeñas que sea factible a costa de reducir las tolerancias de las resistencias R_1 y R_2. El modo común de la salida se maximiza cuanto mayores sean las discrepancias y el peor caso será cuando mayores sean ambas.

En el caso del amplificador diferencial, el efecto de las derivas térmicas de los valores de las resistencias se traducirá en un cambio de ganancia pero no cabe pensar en discrepancias entre los valores que deben ser iguales puesto que dada la proximidad física de todos los componentes cabe esperar que estén sometidos a la misma temperatura y, si como es lógico, son del mismo tipo, evolucionarán de la misma forma.

EJERCICIO 3.9

Considérese el amplificador diferencial de la figura 3.31 en el que se han usado resistencias de R_1 = 1K y R_2 = 10K del 5% de tolerancia para conseguir un amplificador de ganancia 10. Determine el efecto de las variaciones de valor en las resistencias en el caso más desfavorable. ¿Qué ocurriría si se utilizasen resistencias de precisión del 0,1% de tolerancia?

SOLUCIÓN:

El caso más desfavorable es cuando los pares de resistencias sean de los valores más dispares que sea posible, es decir, las de la rama no inversora las consideraremos variando –5%, mientras que las de la rama inversora variarán +5% con lo que los valores que nos encontraremos serán los que se muestran en la figura 3.32.

La expresión de la tensión de salida será, según la expresión (2.49):

$$v_o = \frac{10500}{1050}(v_2 - v_1) + \frac{9500 \cdot 100 + 950 \cdot 1000}{(950 + 9500) \cdot 1050} v_2 = 10(v_2 - v_1) + 0{,}173 v_2$$

apareciendo a la salida más del 17% de v_2. Como se puede ver, el efecto de las discrepancias entre resistencias es muy superior al de la CMRR a baja frecuencia por lo que habrá que tener un especial cuidado en la selección de las resistencias.

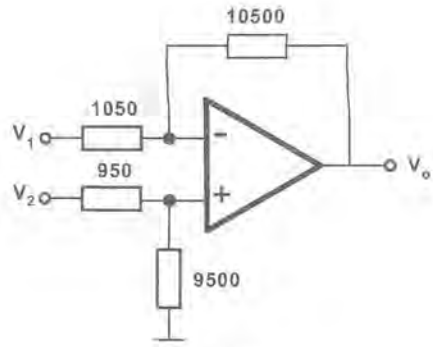


Figura 3.32

En el otro caso, usando resistencias de precisión, tendríamos como peor caso cuando las discrepancias fuesen de 0,1%, es decir, las resistencias de valor 1K podrían llegar a ser de 1001 y 999 Ω y las de 10K, de 10010 ó 9990 Ω. En esas circunstancias:

$$v_o = \frac{10010}{1001}(v_2 - v_1) + \frac{9990 \cdot 1 + 999 \cdot 10}{(999 + 9990) \cdot 1001} v_2 = 10(v_2 - v_1) + 0{,}0018164 v_2$$

con un efecto muy inferior al del caso anterior.

3.6.2 CIRCUITOS INTEGRADORES

El integrador de la tabla 2.1 es uno de los casos de circuito analógico que sólo funciona en teoría puesto que, en la práctica, las no idealidades del operacional provocarán la saturación inmediata de la tensión de salida.

En la figura 3.33a se muestra el integrador con un operacional que posee tensión de desviación, mientras que en la figura 3.33b se muestra el integrador con un operacional en que aparecen las corrientes de polarización.

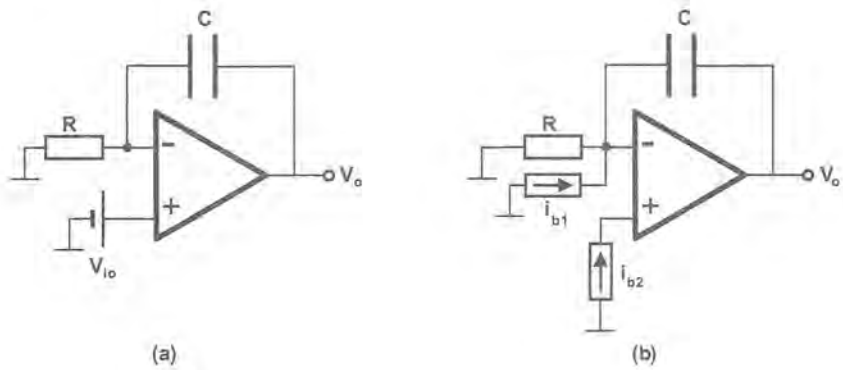


Figura 3.33. a) Integrador con tensión de desviación; b) integrador con corrientes de polarización

En el caso de la tensión de desviación, hará circular una corriente por la resistencia R que provocará la carga del condensador C hasta llegar a la saturación del dispositivo puesto que el sistema tiene ganancia

infinita en continua. Las corrientes de polarización provocarán la aparición de una tensión en la resistencia de entrada de valor $R \cdot i_b$ que se transmitirá a la salida como un error en el cálculo. Por lo tanto, considerando los dos efectos unidos, el circuito no podrá funcionar correctamente a menos que se ajuste la tensión de desviación. Aun en este caso, cualquier deriva térmica producirá un cambio en el valor ajustado y el sistema integrará el error hasta llevarlo a la saturación.

Para resolver el problema existen dos alternativas:

a) Considerar la posibilidad de tener el condensador cortocircuitado siempre salvo cuando se vaya a utilizar el circuito y la tensión de desviación ajustada a 0 V de tal forma que, durante un breve espacio de tiempo se dispondrá del circuito funcionando como está previsto, es decir, integrando la tensión de entrada (figura 3.34).

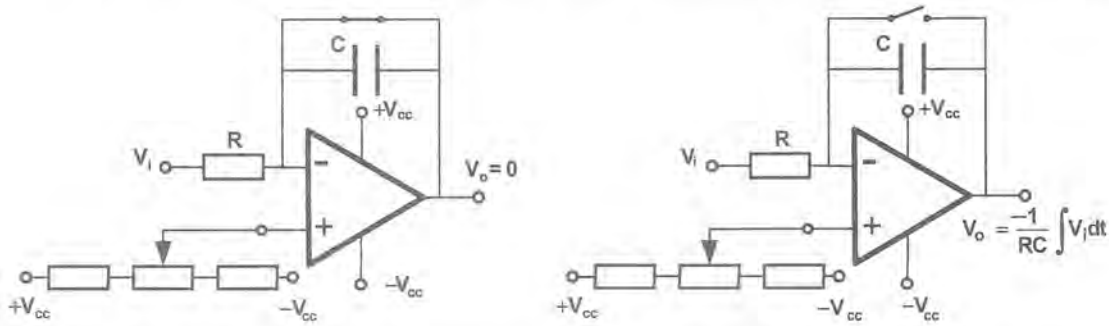


Figura 3.34. Técnica del condensador en cortocircuito.

El tiempo que durará el funcionamiento no puede predecirse *a priori* puesto que dependerá de los cambios térmicos que se produzcan pero nunca suele ser demasiado. Este tipo de técnicas se usaron durante algún tiempo en los computadores analógicos, hoy en desuso.

b) La otra posibilidad es la introducción de una resistencia en paralelo con el condensador C que establezca una ganancia en continua alta pero no infinita. De esta forma, el sistema no integrará los valores de continua ya que, para ellos, presentará una ganancia determinada por las dos resistencias pero tampoco integrará la tensión de desviación que tan sólo será amplificada con lo que el circuito funcionará pero introducirá un error en la salida (figura 3.35).

En este caso, la tensión de desviación introducirá en la salida un valor de continua definido por la ganancia del correspondiente amplificador no inversor, es decir:

$$v_{od} = \frac{R + R'}{R} v_d \tag{3.21}$$

y la corriente de polarización introducirá una tensión de salida similar a la del caso anterior, excepto que ahora la resistencia sobre la que circula esa corriente será el paralelo de R y R'. El circuito no funcionará como integrador en continua sino a partir de una frecuencia f_{min} determinada por R' y C que resulta ser:

$$f_{min} = \frac{1}{2\pi R'C} \tag{3.22}$$

es decir, a medida que R' sea más grande, el circuito es capaz de integrar desde menores frecuencias pero, como contrapartida, según la expresión (3.21), introducirá más errores en continua.

En cualquier caso, el uso de los integradores en circuitos electrónicos está disminuyendo puesto que los sistemas de cálculo digital basados en microprocesadores, microcontroladores o DSPs ofrecen más versatilidad y mucha mayor precisión.

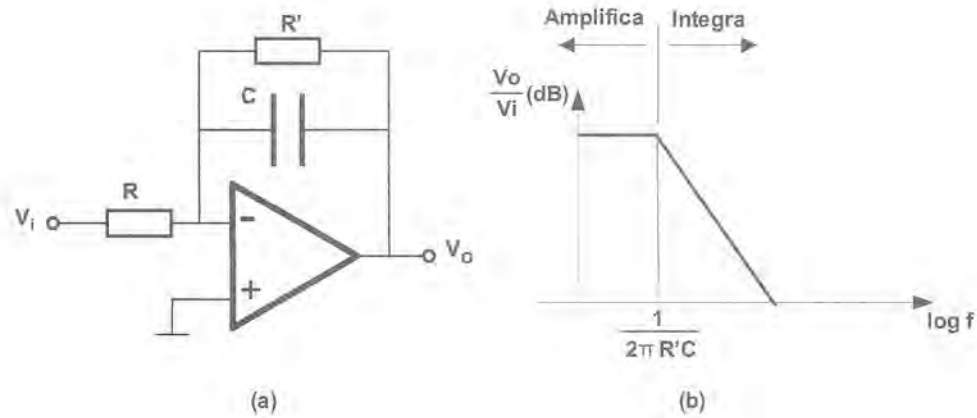


Figura 3.35. a) Integrador con circuito R'C; b) función de transferencia.

3.6.3 CIRCUITOS DERIVADORES

Los circuitos derivadores como el que se muestra en la tabla 2.1 tampoco funcionan bien, presentando algunos inconvenientes producidos por su respuesta en frecuencia. El factor de realimentación $1/\beta$, del circuito derivador resulta ser (figura 3.36a):

$$1/\beta = 1 + j\omega RC$$

que, si se representa junto con la ganancia en lazo abierto del operacional, ocasionará el corte de las dos curvas con pendiente relativa de 40 dB/déc lo que implica un claro peligro de oscilación (figura 3.36b).

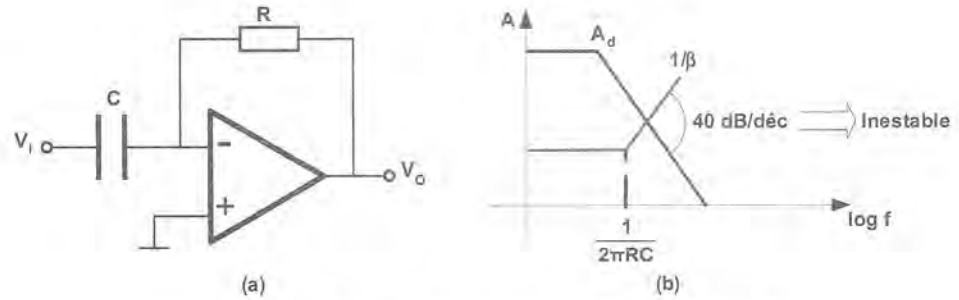


Figura 3.36. a) Derivador; b) función de transferencia.

La forma de paliar el problema es situar una segunda resistencia R' en paralelo con el condensador, de tal manera que haga que $1/\beta$ sea horizontal en alta frecuencia y que corte a la ganancia en lazo abierto con pendiente relativa de 20 dB/déc.

Para ello, se debe incluir una resistencia (figura 3.37a) cuyo valor sea suficientemente grande:

$$R' > \frac{1}{2\pi C f_0}$$

El sistema será entonces estable ya que la curva $1/\beta$ cortará a la ganancia en lazo abierto del operacional con pendiente relativa de 20 dB/déc según se muestra en la figura 3.37b.

El valor de la frecuencia máxima $f_{máx}$ a la que el circuito derivador funcionará como tal vendrá determinada por esa resistencia y por el condensador C, de tal manera que:

$$f_{máx} = \frac{1}{2\pi R'C} \tag{3.23}$$

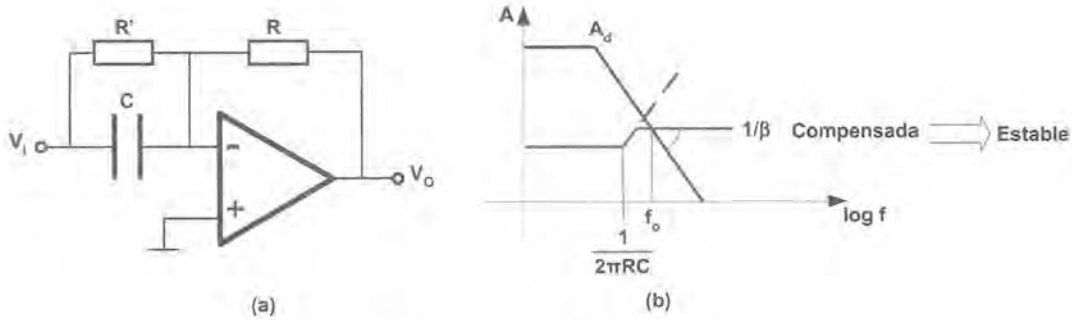


Figura 3.37. a) Derivador con red R'C; b) función de transferencia.

Resumen

- La elección de un determinado tipo de amplificador operacional para un circuito amplificador tiene que responder a una rigurosa caracterización del circuito centrada en el estudio minucioso de las características de la fuente de señal y la carga, la alimentación DC disponible, el consumo permitido y las exigencias de calidad que el circuito tenga, tales como relación señal-ruido a la salida, ganancia, ancho de banda, distorsión de frecuencia y fase y estabilidad térmica de estos parámetros.
- La integración en una pastilla de un conjunto de operacionales, resistencias y otros elementos con el propósito de reducir volumen de circuito, consumo de energía, así como los errores producidos por el desequilibrio de los valores de componentes, el ruido y los efectos de los cambios de temperatura, hacen de los circuitos amplificadores de instrumentación una de las mejores opciones para la amplificación de señales de sensores. Particular importancia poseen los amplificadores aislados al posibilitar un importante aislamiento entre los circuitos de entrada y salida del amplificador.
- El uso de amplificadores operacionales de propósito especial se encuentra estrechamente ligado a aplicaciones en las que el uso de operacionales de propósito general no ofrece una solución adecuada a las exigencias técnico-económicas del circuito.
- Estas consideraciones que hacen referencia a las no-idealidades de los componentes son sólo un ejemplo de los aspectos que deben ser tenidos en cuenta a la hora de abordar el diseño de un circuito que utilice amplificadores operacionales. Cualquier circuito, sea el que sea, no deberá ser tratado como ideal salvo que, después de haber tenido en cuenta todos los aspectos mencionados, los errores o deficiencias obtenidos sean lo suficientemente pequeños como para ser despreciados para la aplicación de que se trate.
- En el campo de la instrumentación electrónica los errores obtenidos y las divergencias respecto del comportamiento ideal hacen que, en la gran mayoría de las aplicaciones, los circuitos con operacionales deban ser analizados con todo rigor y tenido en cuenta casi todas sus no-idealidades. Si no lo hacemos así, tendremos sorpresas en su funcionamiento, ¡y no serán agradables!

Problemas propuestos

3.1 Suponiendo un amplificador diferencial de ganancia 100 realizado con resistencias de 500 Ω y 50K del 5% de tolerancia y un operacional ideal, determine el valor de la CMRR del circuito.

Si la entrada es de 100 mV y aparece un ruido de modo común en la entrada debido a la red eléctrica de 2 V de pico y 50 Hz, ¿cuál es el valor de la relación S/N a la salida del circuito? ¿Cómo se modificaría esa relación si la tolerancia de las resistencias fuese del 0,1%?

3.2 Diseñe un amplificador diferencial con el operacional LF351 que reúna los siguientes requisitos:

Ad = 10
CMRR = 60 dB

¿Cómo se ven afectados estos parámetros si la resistencia de la fuente es de 250 Ω, y totalmente desbalanceada?

3.3 Diseñe un amplificador con el amplificador de instrumentación INA126 capaz de ofrecer una salida en el margen de 0 ÷ 10 V para una entrada de 0 ÷ 50 mV. Si las fuentes de alimentación son de ±15 V; calcule: a) el valor de R_G; b) el valor de V_{iCM} máximo para que funcione sin distorsión de amplitud en la salida del circuito; c) la frecuencia que puede llegar a tener la fuente de señal sin que el amplificador introduzca una apreciable distorsión de frecuencia en la salida. Utilice la documentación del fabricante que se encuentra en el CD.

3.4 En el amplificador SPGIA de la figura 3.7, R_1 = 20 kΩ y R = 10 kΩ. Considere que las resistencias R_{ON} del multiselector son despreciables. Calcule los valores de R_2, R_3 y R_4 si se desea que la tabla de asociación código-ganancia sea la siguiente:

A_0	A_1	Ad
0	0	1
1	0	12
0	1	62
1	1	104

Si el multiselector utilizado posee R_{ON} = 55 Ω, ¿cómo influye esta no-idealidad en los valores de las ganancias?

3.5 A partir de operacionales ideales de propósito general diseñe un amplificador de transconductancia para lograr una salida de 0...10 mA ante una entrada de 0...10 V. La resistencia de carga máxima es R_L = 500 Ω. Utilice la configuración presentada en la tabla 2.1.

3.6 Diseñe un amplificador de tensión en configuración no-inversora sobre la base del amplificador de transimpedancia THS3001 cuya hoja de características se ofrece en el CD. Las fuentes DC disponibles son de ±15 V y se requiere un ancho de banda a pequeña señal (–3 dB) de 100 MHz. Si R_F = 560 Ω, y R_L = 150 Ω con V_o = 300 mVRMS, ¿cuál será la mínima señal de entrada que logra estas condiciones en el circuito?

4

Filtros analógicos

4.1 Introducción
4.2 Clasificación de filtros
4.3 Especificaciones de un filtro
4.4 Aproximaciones matemáticas
4.5 Filtros pasivos RC
4.6 Filtros activos
4.7 Filtros activos de variables de estado
4.8 Filtros de capacidad conmutada
4.9 Selección de componentes
4.10 Programas de ayuda al diseño de filtros
4.11 Comparación de tecnologías de filtros
4.12 Aplicaciones de los filtros

4.1 Introducción

Este capítulo está centrado en el diseño de los filtros analógicos más habituales en el campo de la instrumentación electrónica. Esta labor no siempre es una tarea fácil debido a la variedad de tipos, configuraciones, tecnologías y terminología empleada. A fin de simplificar la exposición se ha prescindido de los desarrollos matemáticos que acompañan a la teoría de filtros y que pueden encontrarse en las publicaciones referenciadas.

Los filtros encuentran aplicación en numerosos campos de la electrónica, destacando dos especialmente: el de las telecomunicaciones (TV, radio, centrales telefónicas, módem, etc.) y el de la instrumentación electrónica (instrumentación en general, sistemas de adquisición de datos, procesamiento de señales, etc.). En el capítulo 1 hablábamos del filtrado como una de las funciones básicas dentro de la cadena de procesamiento de una señal analógica.

En general, un filtro eléctrico puede definirse como un cuadripolo capaz de atenuar determinadas frecuencias del espectro de la señal y permitir el paso sin atenuar de las demás. La figura 4.1 muestra este concepto: la componente del espectro no deseada, de frecuencia f_2, al pasar por el filtro se ve atenuada, mientras que la frecuencia deseada f_1 mantiene su amplitud.

Como tal cuadripolo, un filtro se caracteriza por una función de transferencia H(s) de la forma:

$$H(s) = \frac{V_o(s)}{V_i(s)} \qquad (4.1)$$

donde $s = \sigma + j\omega$ es una variable compleja. En régimen permanente armónico, $s = j\omega$.

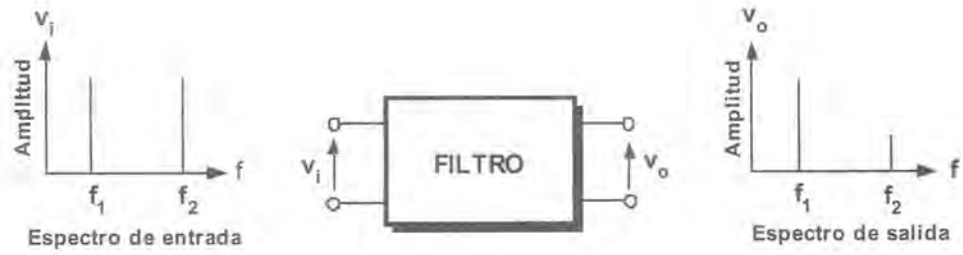


Figura 4.1. Concepto de filtro eléctrico.

La función de transferencia es una expresión compleja y, por lo tanto, podrá expresarse por su módulo o **ganancia** y por su argumento **o fase** de la forma:

$$H(j\omega) = |H(j\omega)| \angle\phi(\omega) \tag{4.2}$$

La ganancia se suele expresar, por comodidad, en decibelios (dB), de la forma

$$20\log|H(j\omega)|$$

Si este valor es negativo, representa una atenuación.

La función de transferencia se puede escribir también como el cociente entre dos polinomios, de la forma:

$$H(s) = \frac{a_m s^m + a_{m-1}s^{m-1} + a_{m-2}s^{m-2} + \cdots + a_1 s + a_0}{s^n + b_{n-1}s^{n-1} + b_{n-2}s^{n-2} + \cdots + b_1 s + b_0} \tag{4.3}$$

Los valores de los coeficientes determinan completamente las características del filtro. El grado del denominador, n, es el **orden del filtro**. Para que el filtro sea realizable, el grado del polinomio denominador debe ser mayor o igual que el grado del polinomio numerador: $n \geq m$.

Otra forma de escribir la función de transferencia de un filtro es factorizar los polinomios numerador y denominador:

$$H(s) = \frac{a_m (s - z_1)(s - z_2)\cdots(s - z_m)}{(s - p_1)(s - p_2)\cdots(s - p_n)} \tag{4.4}$$

Las raíces del denominador, p_1, $p_2 \ldots p_n$, se denominan **polos** de la función de transferencia y las del numerador, z_1, $z_2 \ldots z_m$, **ceros** de la función de transferencia.

4.2 Clasificación de filtros

Los criterios que más se suelen utilizar a la hora de intentar clasificar los distintos tipos de filtros son tres:

1) Según el tipo de señal procesada:

   - **Filtros analógicos:** procesan señales analógicas. Según la tecnología empleada en su realización pueden ser:

- **Filtros pasivos:** diseñados exclusivamente con elementos pasivos R, L, C. Si bien pueden emplearse en cualquier tipo de aplicación, su empleo se restringe a señales de frecuencias elevadas (generalmente superiores a 1 MHz) de forma que la bobina —un componente costoso— resulte pequeña. En potencia se usan en muchas ocasiones; un ejemplo son los filtros que incorporan los altavoces. Por otro lado, no hay que olvidar el filtro RC, de aplicación muy frecuente en electrónica.
- **Filtros activos:** diseñados exclusivamente con resistencias, condensadores y amplificadores operacionales. El empleo de amplificadores operacionales, si bien permite tener una buena adaptación de impedancias entre etapas, condiciona su funcionamiento dinámico al tener limitaciones de ancho de banda.
- **Filtros de capacidad conmutada:** a diferencia de los filtros activos anteriores, la señal de entrada es muestreada y procesada en un tiempo discreto. Las resistencias se reemplazan por interruptores MOS y condensadores integrados en el propio chip. La frecuencia de corte está determinada y es proporcional a la frecuencia de un reloj externo.

- **Filtros digitales:** procesan una señal digital. La señal analógica es cuantificada previamente mediante un convertidor analógico-digital y posteriormente procesada mediante algoritmos software. Al ser programable por software es posible cambiar tanto la ganancia como la fase del filtro sin necesidad de cambiar el hardware.

2) Según la función que lleven a cabo:

- **Filtros paso bajo:** dejan pasar todas las frecuencias por debajo de una determinada frecuencia de corte, f_c, atenuando las frecuencias por encima de dicha frecuencia de corte. La figura 4.2a muestra la curva de la ganancia de esta función en función de la frecuencia. Para $f = f_c$, la ganancia cae 3 dB, con respecto a la ganancia máxima, que por comodidad se ha supuesto unidad (es decir, 0 dB). La frecuencia f_c se denomina frecuencia de corte y su definición es válida para los cuatro tipos de filtros.
- **Filtros paso alto:** atenúan las frecuencias bajas y dejan pasar aquéllas que estén por encima de la frecuencia de corte, f_c (figura 4.2b).
- **Filtros paso banda:** dejan pasar aquellas frecuencias comprendidas entre la frecuencia de corte inferior, f_{c1}, y la frecuencia de corte superior, f_{c2}, atenuando las demás (figura 4.2c). La frecuencia f_0 se denomina frecuencia central.
- **Filtros de rechazo de banda:** atenúan las frecuencias comprendidas entre las frecuencias de corte inferior (f_{c1}) y superior (f_{c2}), dejando pasar las otras (figura 4.2d).

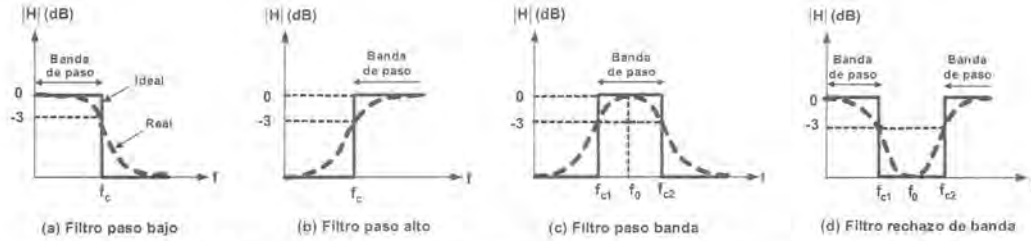


Figura 4.2. Funciones de transferencia ideales y reales de los filtros.

3) Según la aproximación matemática empleada:

Existen diversas funciones matemáticas empleadas para aproximar la respuesta real del filtro a la ideal. Cada una de estas aproximaciones busca optimizar una determinada característica del filtro. Se tienen diversas aproximaciones, que dan nombre a los filtros que las emplean, las más comunes son:

- **Filtros de Butterworth:** el objetivo de la aproximación de Butterworth es obtener una respuesta de ganancia plana en la banda de paso. Esto se consigue a costa de una región de transición de caída lenta y una respuesta de fase no lineal alrededor de la frecuencia de corte.
- **Filtros de Chebyshev:** el objetivo de la aproximación de Chebyshev es maximizar la pendiente de la característica de ganancia en la región de transición. Por el contrario, presenta un cierto rizado en la banda de paso, que se incrementa al aumentar el orden del filtro.
- **Filtros de Bessel:** el objetivo de la aproximación de Bessel es lograr una respuesta de fase lineal en un margen de frecuencias amplio en torno a la frecuencia de corte. La ganancia de la banda de paso no es tan plana como en un filtro de Butterworth ni la pendiente en la banda de transición tan acentuada como en un filtro de Chebyshev.

La figura 4.3 muestra una representación grafica de la función que realiza el filtro. Así, por ejemplo, para un filtro paso bajo o paso alto, la onda senoidal superior que aparece dentro del rectángulo representa la banda de frecuencias por encima de la frecuencia de corte y la inferior la banda por debajo de la frecuencia de corte. Si la onda aparece cortada, indicará que dicha banda de frecuencias estará atenuada.

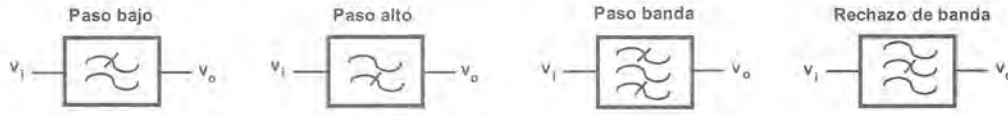


Figura 4.3. Representación gráfica de un filtro según la función que realice.

4.3 Especificaciones de un filtro

Si bien la ganancia del filtro, su orden y la frecuencia de corte son las especificaciones que con más frecuencia se suelen tomar para caracterizar un filtro, dependiendo de la aplicación puede ser necesario especificar otras características, tanto en el dominio de la frecuencia como en el dominio del tiempo.

4.3.1 ESPECIFICACIONES FRECUENCIALES

En general, un filtro puede ser especificado en el dominio de la frecuencia, fijando alguno de los siguientes parámetros:

- $A_{máx}$ es el máximo cambio permitido en la ganancia dentro de la banda de paso. También se suele denominar rizado máximo de la banda de paso.
- $A_{mín}$ es la atenuación mínima de la banda suprimida (referida a la ganancia máxima de la banda de paso).
- f_p es la frecuencia límite de la banda de paso.
- f_s es la frecuencia a la que comienza la banda suprimida.

Estos parámetros se han representado gráficamente en la figura 4.4 para el caso de un filtro paso bajo y para uno paso banda. Como se aprecia, la respuesta real del filtro ha de estar contenida entre los límites impuestos por dichas especificaciones.

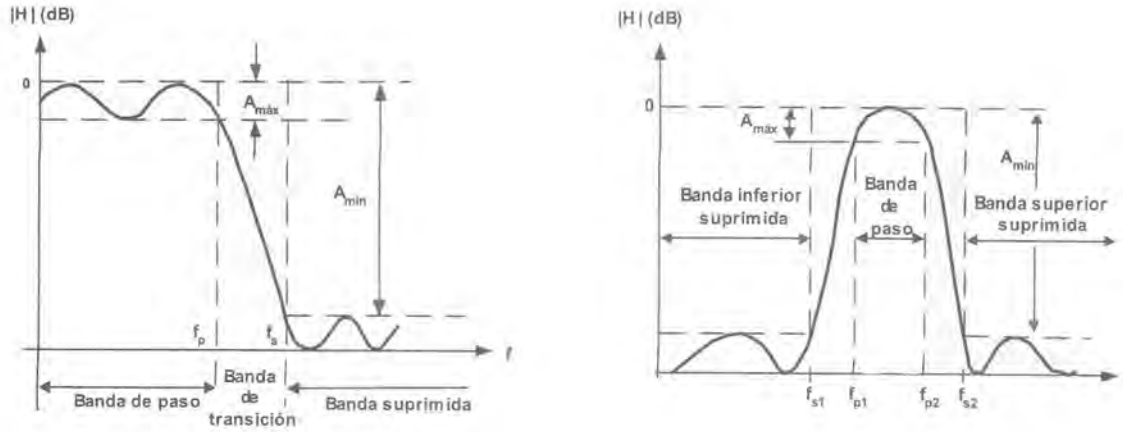


Figura 4.4. Especificaciones típicas de un filtro en el dominio de la frecuencia.

Otro parámetro de diseño es el **factor de calidad del filtro** (Q). Es un parámetro equivalente al orden del filtro. Podemos decir que diseñar un filtro de un orden determinado es equivalente a hacerlo para un valor específico de Q. Para un filtro paso banda se define como la relación entre la frecuencia central, f_0 y el ancho de banda, B:

$$Q = \frac{f_0}{B} \tag{4.5}$$

En los filtros paso bajo y paso alto a medida que Q crece lo hace también el valor del pico de la respuesta en ganancia (figura 4.5a). En el caso de un filtro paso banda o de rechazo de banda se observa cómo el filtro se va haciendo más selectivo a medida que el valor de Q aumenta (figura 4.5b).

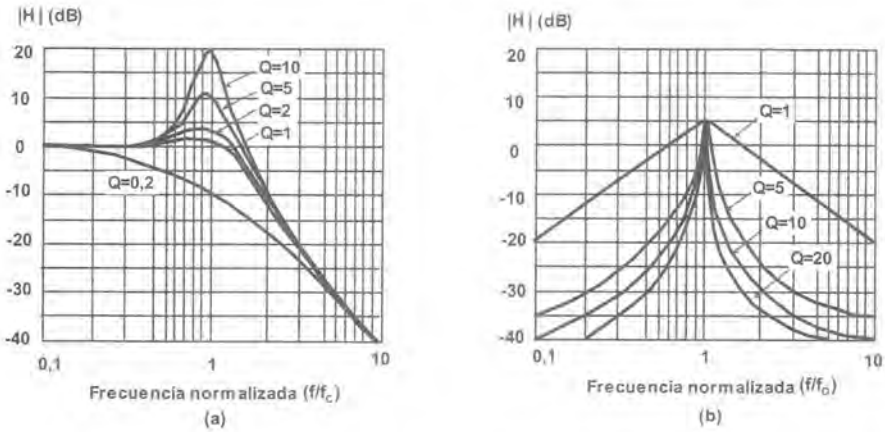


Figura 4.5. Respuestas normalizadas de varios filtros de segundo orden en función de Q.

Respuesta frecuencial normalizada

En adelante, como ya se hizo en la figura 4.5, el eje de frecuencias se representará dividido por una frecuencia característica: la frecuencia de corte, f_c, para los filtros paso bajo y paso alto o por la frecuencia central, f_0, en el caso de los filtros paso banda y de rechazo de banda. De la misma forma la pulsación ω, quedará

normalizada sin más que dividirla por ω_c. Las expresiones de la frecuencia normalizada, f_n y de la pulsación normalizada, ω_n, son, por tanto:

$$f_n = \frac{f}{f_c}; \quad \omega_n = \frac{\omega}{\omega_c} \tag{4.6}$$

Esta forma de proceder aporta las siguientes facilidades en el estudio de los filtros:

a) Permite una rápida y eficaz comparación de los distintos tipos de filtros.
b) Permite un estudio más consistente.
c) Es la base del método expuesto en el apartado 4.6 para el diseño de los filtros activos.

Por otro lado, el eje de frecuencias normalizadas se representará en una **escala logarítmica**, a fin de poder visualizar todo el margen de frecuencias de interés. Por simplicidad, en adelante se entenderá que el eje de frecuencias está a escala logarítmica, aunque no se indique de forma explícita.

4.3.2 ESPECIFICACIONES TEMPORALES

Para caracterizar la respuesta temporal de un filtro se acude a conocer la respuesta del filtro ante una entrada escalón de tensión (*step response*). Los principales parámetros temporales (tiempo de subida, tiempo de establecimiento y sobreoscilación) ya han sido definidos en el capítulo 1. Lo deseable de la respuesta temporal es que la salida alcance el valor final de forma rápida, sin sobreoscilación ni rizado. La figura 4.6 compara las respuestas temporales de las aproximaciones de Butterworth, Bessel y Chebyshev ante una entrada escalón de tensión. La aproximación de Bessel es la que tiene un mejor comportamiento en el dominio del tiempo; a continuación, le sigue Butterworth y en último lugar está Chebyshev que, como vemos, es el que presenta un mayor rizado en la banda de paso, así como sobreoscilación.

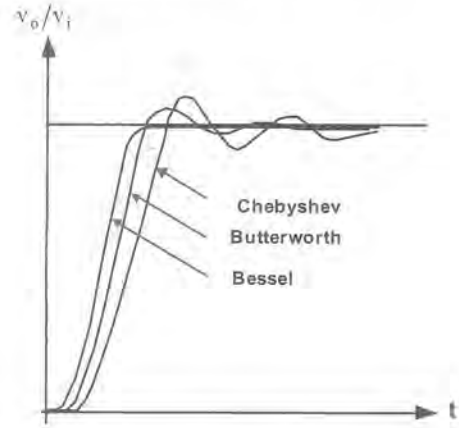


Figura 4.6. Comparación de respuestas temporales a una entrada escalón.

Retardo de fase y retardo de grupo

En los casos donde a la entrada del filtro se aplican señales periódicas complejas[1], es más conveniente hablar de retardo de tiempo que de fase. Para cuantificar la relación entre fase y tiempo supongamos que a la entra-

[1] Toda señal periódica puede representarse como la suma de señales senoidales con frecuencias múltiplos enteros de la frecuencia fundamental.

da de un filtro se aplica una señal senoidal sen(ωt) y en la salida se tiene sen(ωt-ϕ), donde ϕ es el desfase. Al desfase ϕ entre las dos señales le corresponde un retardo de tiempo equivalente, ϕ/ω, conocido como **retardo de fase**. Por ejemplo, una señal de 1 kHz que sufre un desfase de 36° al atravesar el filtro, experimenta un retardo de fase equivalente de 0,1 ms. La representación del retardo de fase en lugar de la fase suele ser, en algunas ocasiones, más conveniente ya que el retardo no cambia tan bruscamente como lo pueda hacer la fase. El término **retardo de grupo** expresa el retardo de fase visto por el grupo de frecuencias que se transmiten a través del filtro, es decir, en la banda de paso. Matemáticamente, el retardo de grupo se define como la derivada de la fase con respecto a la frecuencia:

$$T_D(\omega) = -\frac{d\phi(\omega)}{d\omega} \tag{4.7}$$

La figura 4.7 muestra la respuesta de fase, y el retardo de grupo de las distintas aproximaciones matemáticas que hemos venido considerando. Los filtros de Bessel tienen un retardo de grupo constante, es decir, introducen un retardo constante para todas las frecuencias dentro de la banda de paso. En cambio, los filtros de Butterworth y Chebyshev tienen un retardo de grupo que depende de la frecuencia y la velocidad de cambio de la fase se incrementa al aproximarse a la frecuencia de corte.

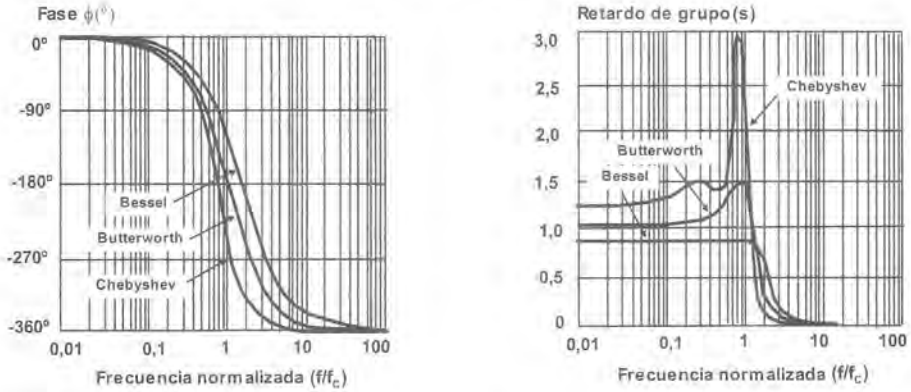


Figura 4.7. Comparación de la respuesta de fase y retardos de grupo normalizados de filtros de 4° orden.

Como consecuencia de no tener un retardo de grupo constante, la salida crece más lentamente, se tiene sobreoscilación y rizado (figura 4.6). El retardo de grupo es especialmente importante cuando a la entrada del filtro se aplican pulsos. En este caso, tanto la frecuencia fundamental del pulso como los armónicos experimentan distintos retardos, dando lugar a un rizado en el pulso de salida. Estas distorsiones de la salida del filtro pueden producir errores en las etapas conectadas a continuación.

4.4 Aproximaciones matemáticas

En la figura 4.2 presentábamos las respuestas de ganancia de cada uno de los cuatro tipos de filtros. En este apartado presentaremos las aproximaciones matemáticas más comunes a dichas características. Para ello tomaremos como ejemplo el filtro paso bajo, aunque es posible extender el estudio a otro tipo de filtro sin más que realizar las transformaciones frecuenciales adecuadas.

Aproximación de Butterworth

La respuesta de ganancia de un filtro paso bajo de Butterworth viene dada por la expresión matemática:

$$|H(jf)| = \frac{1}{\sqrt{1+\epsilon^2\left(\frac{f}{f_p}\right)^{2n}}} \tag{4.8}$$

donde n es el orden del filtro, ϵ es el parámetro que determina el rizado máximo de la banda de paso, $A_{máx}$ y f_p es la frecuencia extrema de la banda de paso.

La expresión 4.8 se ha representado en la figura 4.8 para varios valores de n. Se observa que es una curva monótona decreciente en todas las frecuencias.

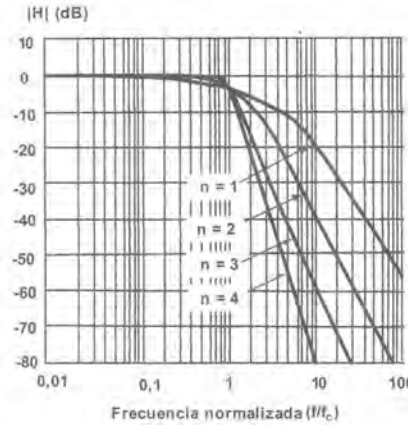


Figura 4.8. Respuesta de ganancia para filtros Butterworth de diversos órdenes.

Para $f = f_p$ se tiene el rizado máximo de la banda de paso, $A_{máx}$, es decir:

$$|H(jf_p)| = \frac{1}{\sqrt{1+\epsilon^2}} = A_{máx} \tag{4.9}$$

Por tanto, dado el rizado máximo de la banda de paso, $A_{máx}$ (suele estar entre 0,5 dB y 3 dB), el valor del parámetro ϵ se obtiene mediante:

$$\epsilon = \sqrt{10^{A_{máx}/10} - 1} \tag{4.10}$$

En el borde de la banda suprimida, es decir, para $f = f_s$ se tiene la atenuación mínima del filtro, A_{min}, por lo tanto:

$$|H(jf_s)| = \frac{1}{\sqrt{1+\epsilon^2\left(\frac{f_s}{f_p}\right)^{2n}}} = A_{min} \tag{4.11}$$

A partir de esta ecuación se puede determinar el orden mínimo del filtro que satisface las especificaciones de un rizado máximo en la banda de paso, $A_{máx}$ y una atenuación mínima, A_{min} en la banda suprimida. Operando se obtiene:

$$n = \frac{\log\left[\left(\frac{1}{A_{min}^2}\right) - 1\right] - 2 \cdot \log \in}{2 \cdot \log\left(\frac{f_s}{f_p}\right)} \qquad (4.12)$$

En esta expresión A_{min} no está expresada en dB.

EJERCICIO 4.1

Para transmitir una señal de voz digitalmente a través de un canal telefónico se requiere un filtro paso bajo con una banda de paso de 3000 Hz y una banda de transición de aproximadamente 500 Hz. Determine el orden del filtro de Butterworth necesario si se requiere una atenuación mínima de 25 dB en la frecuencia de 3500 Hz. Suponga un rizado máximo en la banda de paso de 3 dB.

SOLUCIÓN:

Las especificaciones del filtro se han representado en la figura 4.9. La respuesta del filtro ha de estar dentro del área rayada. Sustituyendo $A_{máx} = 3$ dB en la ecuación (4.10) se obtiene el valor del parámetro $\in$ que determina el rizado máximo en la banda de paso, $A_{máx}$.

$$\in = \sqrt{10^{A_{máx}/10} - 1} = \sqrt{10^{3/10} - 1} = 0{,}9976 \cong 1$$

Una atenuación mínima de 25 dB equivale a una $A_{min} = 0{,}056$. Aplicando la ecuación (4.12) con: $\in = 1$, $f_p = 3000$ Hz, $f_s = 3500$ Hz y $A_{min} = 0{,}056$ resulta un orden $n = 19$. Compruebe cómo relajando un poco las especificaciones del filtro, por ejemplo, tomando $A_{min} = 20$ dB, el orden del filtro se reduce a 8.

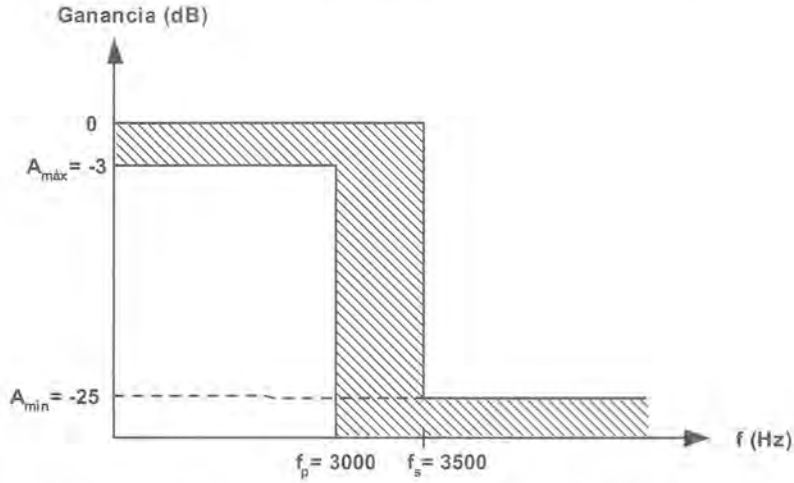
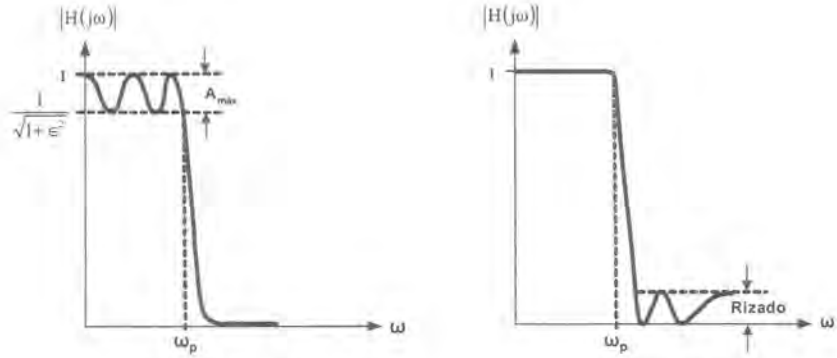


Figura 4.9. Especificaciones de la respuesta de ganancia del ejercicio 4.1.

Aproximación de Chebyshev

Hay dos aproximaciones de Chebyshev diferentes: la primera da lugar a los denominados "**filtros tipo I**" que sólo tienen polos y presentan un rizado constante en la banda de paso y una característica monótona en la banda suprimida (figura 4.10a). La segunda aproximación da lugar a los "**filtros de tipo II**" que contienen polos y ceros. Se caracterizan por un comportamiento monótono en la banda de paso y un comportamiento de rizado constante en la banda suprimida, estando los ceros en el eje imaginario del plano s (figura 4.10b).



(a) Chebyshev Tipo I (b) Chebyshev Tipo II

Figura 4.10. Características del filtro de Chebyshev.

El módulo de la función de transferencia de un filtro de Chebyshev de tipo I viene dada por:

$$|H(j\omega)| = \frac{1}{\sqrt{1+\epsilon^2 \cos^2\left[n \cdot \cos^{-1}\left(\omega/\omega_p\right)\right]}} \quad \text{para } \omega \le \omega_p \tag{4.13}$$

$$|H(j\omega)| = \frac{1}{\sqrt{1+\epsilon^2 \cosh^2\left[n \cdot \cosh^{-1}\left(\omega/\omega_p\right)\right]}} \quad \text{para } \omega > \omega_p \tag{4.14}$$

Para $\omega = \omega_p$, resulta:

$$|H(j\omega_c)| = \frac{1}{\sqrt{1+\epsilon^2}} = A_{máx} \tag{4.15}$$

El valor del parámetro ϵ, al igual que para la aproximación de Butterworth, se calcula por la expresión (4.10).

Para un conjunto dado de especificaciones ϵ, A_{min}, ω_s y ω_p, el orden del filtro lo podemos determinar a partir de la expresión:

$$n = \frac{\log\left[\left(\sqrt{1-A_{min}^2} + \sqrt{1-A_{min}^2\left(1+\epsilon^2\right)}\right)/\epsilon \cdot A_{min}\right]}{\log\left[\left(\omega_s/\omega_p\right) + \sqrt{\left(\omega_s/\omega_p\right)^2 - 1}\right]} \tag{4.16}$$

EJERCICIO 4.2

Determine el orden de un filtro de Chebyshev de tipo I que satisfaga las siguientes especificaciones: rizado máximo en la banda de paso 1 dB, banda de paso de 3000 Hz, frecuencia extrema de la banda suprimida 3500 Hz, y atenuación mínima de 20 dB para $\omega \geq \omega_s$.

SOLUCIÓN:

El parámetro del filtro ϵ se obtiene de la expresión (4.10), donde el valor de $A_{max} = 1$ dB.

$$\epsilon = \sqrt{10^{A_{max}/10} - 1} = \sqrt{10^{1/10} - 1} = 0,5088$$

Entrando en la expresión (4.16), con $\epsilon = 0,5088$, $f_p = 3000$ Hz, $f_s = 3500$ Hz y $A_{min} = 0,1$ se obtiene el orden del filtro de Chebyshev tipo I:

$$n = \frac{\log\left[\left(\sqrt{1-0,1^2} + \sqrt{1-0,1^2\left(1+0,5088^2\right)}\right)/0,5088 \cdot 0,1\right]}{\log\left(1,166 + \sqrt{1,53}\right)} = \frac{\log 39,08}{\log 2,40} = 4,18$$

Tomaremos para n el valor entero más próximo por encima, es decir, $n = 5$.

Aproximación de Bessel

Los filtros basados en la aproximación de Bessel se caracterizan por tener una función de transferencia de la forma:

$$H(s) = \frac{1}{B_n(s)} \tag{4.17}$$

donde $B_n(s)$ es el polinomio de Bessel de n-ésimo orden, que se puede expresar como:

$$B_n(s) = \sum_{k=0}^{n} a_k s^k \tag{4.18}$$

donde los coeficientes $\{a_k\}$ vienen dados por:

$$a_k = \frac{(2n-k)\ !}{2^{n-k}k!(n-k)\ !} \quad k = 0,1,\ldots,n \tag{4.19}$$

Una característica importante de los filtros de Bessel es que su respuesta de fase es lineal en la banda de paso del filtro, lo cual da lugar a un retardo de grupo constante en ese margen de frecuencias. Por tanto, es una buena elección en aplicaciones de filtrado de ondas cuadradas.

Comparación de aproximaciones

Una vez presentadas las aproximaciones matemáticas más comunes, estamos en condiciones de poder realizar un análisis comparativo entre todas ellas. La figura 4.11 compara la respuesta de ganancia de filtros paso bajo de cuarto orden. A partir de la información proporcionada por las figuras 4.6, 4.7 y 4.11 se ha completado la tabla 4.1 que resume las ventajas e inconvenientes de las tres aproximaciones analizadas. Una aproximación no es buena ni mala en sí misma, depende de la aplicación donde se utilice. En términos gene-

rales podemos decir que la aproximación de Butterworth es la que presenta un mejor equilibrio de comportamiento, al presentar una respuesta de ganancia plana en la banda de paso y una linealidad de fase intermedia entre las aproximaciones de Chebyshev y Bessel, dando lugar a una moderada sobreoscilación y rizado en la respuesta a un escalón.

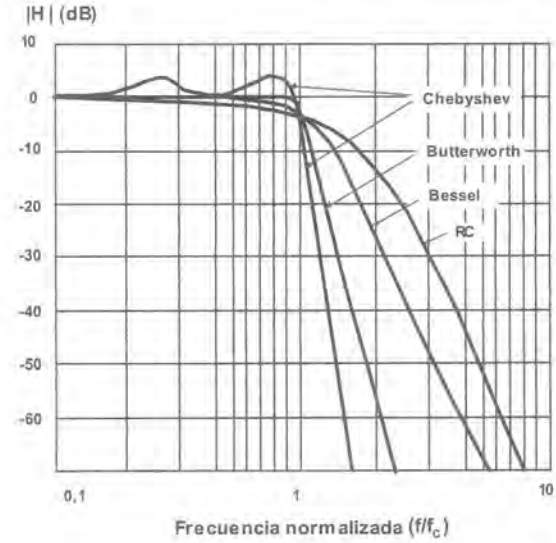


Figura 4.11. Respuestas de ganancia de un filtro paso bajo de 4º orden.

Butterworth	Chebyshev (tipo I)	Bessel
Ventajas: • Respuesta de ganancia plana en la banda de paso. • Respuesta a escalón mejor que Chebyshev. • Atenuación mayor que Bessel por encima de la banda de paso.	Ventajas: • Atenuación mayor que Butterworth por encima de la banda de paso.	Ventajas: • La mejor respuesta a un escalón de tensión.
Inconvenientes: • Un poco de sobreoscilación y rizado en la respuesta a un escalón.	Inconvenientes: • Rizado en la banda de paso. • Respuesta a un escalón peor que Butterworth.	Inconvenientes: • Atenuación más lenta que Butterworth, por encima de la banda de paso.

Tabla 4.1. Comparación de las aproximaciones matemáticas de los filtros.

4.5 Filtros pasivos RC

Los filtros pasivos emplean, como su nombre indica, exclusivamente componentes pasivos (resistencias, bobinas y condensadores). Las principales ventajas que presentan son: buena linealidad, no requieren alimentación y permiten un amplio margen de tensiones y corrientes. Por el contrario, tienen poca flexibilidad

en el diseño, cada etapa tiene una impedancia de salida que afecta a la siguiente y una impedancia de entrada baja que afecta a la anterior, lo que dificulta su conexión en cascada. El tamaño de las bobinas para frecuencias bajas resulta elevado[2], encareciendo el filtro. Además, al estar formados sólo por elementos pasivos la ganancia, en la banda de paso es siempre la unidad.

De todos los filtros pasivos el más sencillo y uno de los más utilizados en instrumentación es el filtro RC, por lo que será al que le prestemos una mayor atención. La figura 4.12 muestra el esquema de un filtro RC de primer orden. Su función de transferencia es:

$$H(j\omega) = \frac{1}{1 + j\omega RC} \tag{4.20}$$

Para poder representar la respuesta frecuencial del filtro (diagrama de Bode), calculemos el módulo y la fase de la función de transferencia:

- Módulo: $|H| = \dfrac{1}{\sqrt{1+\left(\dfrac{f}{f_c}\right)^2}}$ donde $f_c = 1/2\pi RC$

- Fase: $\phi\,(f) = -\text{arctg}\dfrac{f}{f_c}$

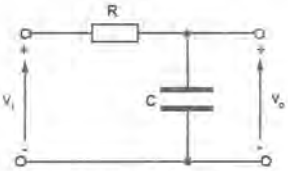


Figura 4.12. Filtro RC de paso bajo.

A fin de simplificar la representación gráfica, es usual realizar una representación asintótica de la función de transferencia. Para ello se consideran tres intervalos:

$f/f_c << 1$: $|H| \to 1$ (0 dB) ; $\phi \to 0$

$f/f_c >> 1$: $|H| \to f_c/f$ ($20\log f_c/f$) ; $\phi \to -\pi/2$

$f/f_c = 1$: $|H| = 0,707$ (−3 dB) ; $\phi = -\pi/4$

A partir de estos valores hemos representado el diagrama de Bode de la figura 4.13. Vemos que se trata de un filtro que realiza la función paso bajo. Para $f/f_c = 1$, es decir, para $f = f_c = 1/2\pi RC$, la ganancia cae 3 dB. Esta frecuencia se conoce como **frecuencia de corte del filtro**. Para $f/f_c >> 1$ su respuesta en ganancia tiene una pendiente de −20 dB/década, con lo que se trata de un filtro de primer orden.

[2] Las bobinas no suelen ser componentes comerciales y es preciso hacerlas para cada aplicación. Esto hace que su coste sea, relativamente, mucho mayor que el de otros componentes reactivos.

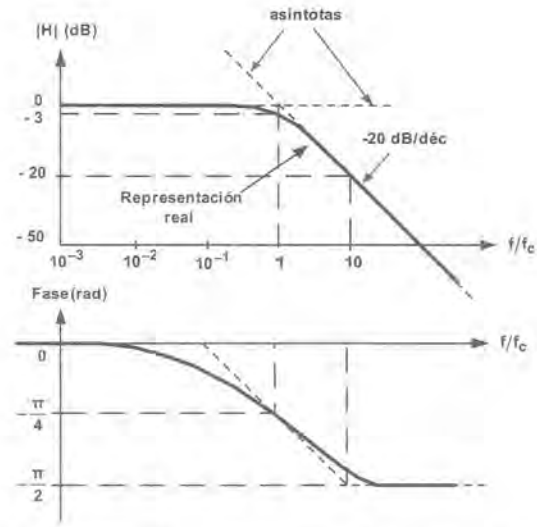


Figura 4.13. Diagrama de Bode de un filtro RC paso bajo de primer orden.

Si queremos aumentar el orden del filtro se pueden conectar en cascada varias etapas RC. Para evitar el efecto de carga entre las etapas pueden colocarse operacionales como seguidores de tensión entre etapas RC, como se muestra en la figura 4.14.

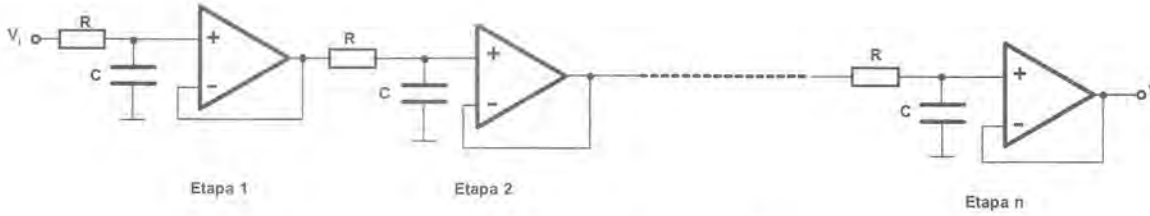


Figura 4.14. Filtro RC de orden n.

Suponiendo ahora que cada etapa representa una carga despreciable para la anterior, la función de transferencia de las n etapas iguales en cascada es:

$$H(jf) = \frac{1}{\left(1 + j\frac{f}{f_c}\right)^n} \tag{4.21}$$

donde f_c es la frecuencia de corte de cada etapa individual. La frecuencia de corte de las n etapas en cascada, f_c', se obtiene como se sabe, igualando el módulo de la función de transferencia al valor $1/\sqrt{2}$ (–3 dB), es decir:

$$|H| = \frac{1}{\left[\sqrt{1 + \left(\frac{f_c'}{f_c}\right)^2}\right]^n} = \frac{1}{\sqrt{2}} \tag{4.22}$$

Operando resulta:

$$f_c' = f_c \cdot \sqrt{2^{1/n} - 1} \tag{4.23}$$

Por ejemplo, para n = 4, $f_c' = 0{,}43 \cdot f_c$, es decir, la ganancia no se mantiene constante en toda la banda de paso, sino que disminuye un poco al acercarnos a la frecuencia de corte de cada etapa individual, f_c, como se muestra en la figura 4.15. Otra importante limitación del filtro RC es que la respuesta de fase no es lineal, incrementando significativamente la distorsión de la señal de salida.

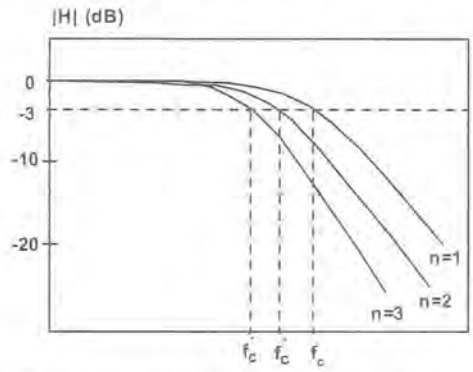


Figura 4.15. Variación de la frecuencia de corte en etapas RC paso bajo en cascada.

Un estudio similar al realizado para el filtro RC paso bajo se puede repetir para otros circuitos pasivos RC. La tabla 4.2 muestra algunas posibles realizaciones. El filtro paso banda se obtuvo conectando en cascada un filtro paso bajo y uno paso alto. Debido al efecto de carga entre las dos etapas la función de transferencia no se puede obtener como el producto de las funciones de transferencia individuales. Se trata por tanto de un filtro de segundo orden con una pendiente de ±20 dB/déc. El filtro de rechazo de banda puede implementarse de múltiples formas, la tabla 4.2 muestra el caso de un filtro pasivo en doble T que rechaza la frecuencia central $f_0 = 1/2\pi RC$. La forma en cuña que tiene su respuesta de ganancia, hace que se denominen **filtros *notch***.

EJERCICIO 4.3

Se pretende leer una señal de 1 Hz, que incluye una interferencia de 1 kHz de su misma amplitud A. Si se pretende atenuar 100 veces la interferencia garantizando que la señal buscada no esté atenuada en más de un 5%. ¿Qué valor debería tener el producto RC (constante de tiempo) de un filtro RC de primer orden que satisfaga las condiciones expuestas?

SOLUCIÓN:

Figura 4.16

De la tabla 4.2, tenemos que para un filtro RC paso bajo de primer orden, la función de transferencia es:

$$H(j\omega) = \frac{1}{1 + j\omega RC}$$

Para f = 1 Hz, es decir, $\omega = 2\pi \cdot 1$ rad/s, la atenuación tiene que ser menor del 5%, o sea:

$$\frac{1}{\sqrt{1+(2\pi \cdot RC)^2}} \geq 0{,}95 \Rightarrow RC \leq 0{,}052 \text{ s}$$

Para f = 1 KHz, la atenuación debe ser mayor que 100, por lo que, para $\omega = 2\pi \cdot 1000$ rad/s, se tiene:

$$\frac{1}{\sqrt{1+(2\pi \cdot 1000 \cdot RC)^2}} \leq 0{,}01 \Rightarrow RC \geq 0{,}016 \text{ s}$$

Los límites de la constante de tiempo son: $0{,}016 \text{ s} \leq \tau \leq 0{,}052 \text{ s}$

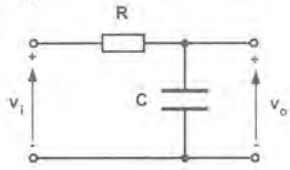
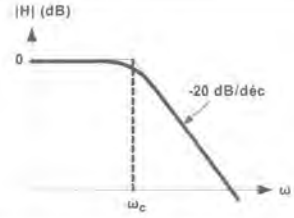
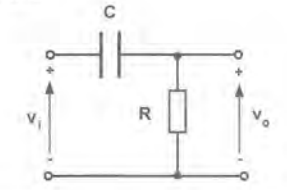
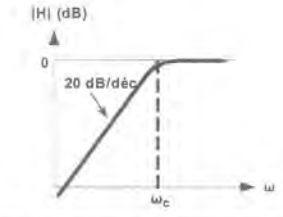
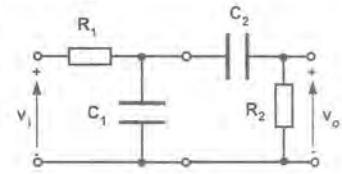
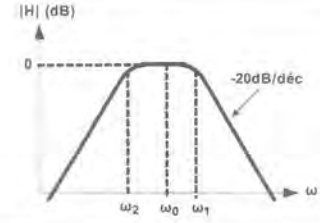
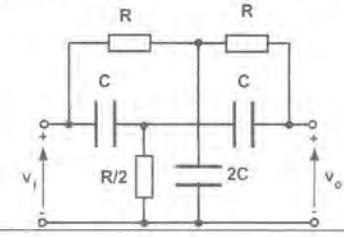
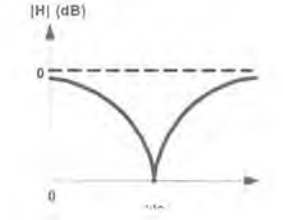
Tipo de filtro	Respuesta frecuencial	Relaciones de interés
Paso bajo 		$H(j\omega) = \frac{1}{1+j\omega RC}$ Frecuencia de corte: $f_c = 1/2\pi RC$ $\omega_c = 2\pi f_c$ Frecuencia de corte de n etapas iguales en cascada y aisladas: $f_c' = f_c \cdot \sqrt{2^{1/n}-1}$
Paso alto 		$H(j\omega) = \frac{j\omega RC}{1+j\omega RC}$ Frecuencia de corte: $f_c = 1/2\pi RC$ $\omega_c = 2\pi f_c$ Frecuencia de corte de n etapas iguales en cascada y aisladas: $f_c' = f_c / \sqrt{2^{1/n}-1}$
Paso banda 		$H(j\omega) = \frac{j\omega\tau_2}{1+j\omega(\tau_1+\tau_2+a\tau_2)-\omega^2\tau_1\tau_2}$ $\tau_1 = R_1C_1;\ \tau_2 = R_2C_2;\ a = R_1/R_2$ $\omega_1 = 1/\tau_1;\ \omega_2 = 1/\tau_2$ Pulsación central: $\omega_0 = \sqrt{\omega_1 \cdot \omega_2}$
Rechazo de banda 		$H(j\omega) = \frac{j\beta}{4+j\beta}$ $\beta = \omega/\omega_0 - \omega_0/\omega$ Pulsación central: $\omega_0 = 1/RC$

Tabla 4.2. Ejemplos de filtros pasivos RC.

4.6 Filtros activos

Se ha visto que los filtros pasivos, en aplicaciones de instrumentación electrónica, se reducen a estructuras RC que si bien son sencillas, no satisfacen totalmente los requerimientos que la instrumentación electrónica suele demandar.

Los filtros activos son una solución válida para gran número de aplicaciones de baja frecuencia y de pequeña señal; por otro lado, las más habituales en el campo de la instrumentación electrónica. Los filtros activos utilizan operacionales para conseguir funciones de transferencia como las aproximaciones explicadas de Butterworth, Chebyshev y Bessel aprovechando las grandes ventajas del operacional, aunque limitados también por su comportamiento real. Las ventajas que aportan los filtros activos son:

a) Utilizan sólo resistencias, condensadores y operacionales.
b) Facilitan el encadenamiento de etapas aprovechando la baja impedancia de salida de los operacionales.
c) Proporcionan una gran flexibilidad de diseño permitiendo implementar cualquiera de las aproximaciones matemáticas que hemos visto.

Sin embargo, los filtros activos no están exentos de inconvenientes, generados en la gran mayoría de los casos por las limitaciones del amplificador operacional. Entre estos inconvenientes destacamos:

a) Ancho de banda limitado al del operacional empleado.
b) Necesidad de una o, como es normal, dos fuentes de alimentación.
c) Limitación en los valores de R y C necesarios, por el comportamiento del operacional.
d) Ruido adicional generado por el propio operacional.

4.6.1 ESTRUCTURAS PARA LA REALIZACIÓN DE FILTROS ACTIVOS

Existen numerosas estructuras físicas para la realización de filtros activos. La tabla 4.3 muestra las características más importantes de tres de las estructuras o topologías más comunes: Sallen-Key, realimentación múltiple y variables de estado. Hay que aclarar que una misma estructura, puede utilizarse para implementar funciones y aproximaciones matemáticas diferentes con sólo cambiar el valor de los coeficientes de la función de transferencia. Las estructuras de Sallen-Key y de realimentación múltiple se emplearán en el diseño de filtros paso bajo y paso alto, mientras que la estructura de variables de estado será analizada de forma independiente en el apartado 4.7.

4.6.2 DISEÑO DE FILTROS PASO BAJO

La necesidad de filtrado paso bajo está presente en un gran número de aplicaciones en el campo de la instrumentación electrónica. Así, los filtros paso bajo son siempre una etapa previa a la conversión analógica-digital. En otras ocasiones se emplean para eliminar el ruido y las interferencias de alta frecuencia que, a menudo, tratan de alterar la señal deseada.

Prescindiendo de las demostraciones matemáticas, que se pueden encontrar en las referencias [4.1] y [4.2], el método general de diseño de los filtros activos está basado en considerar la función de transferencia normalizada de un filtro de segundo orden e identificarla con la expresión matemática del filtro particular a diseñar. La expresión general de la función de transferencia de un filtro paso bajo de segundo orden está dada por[3]:

$$H(j\omega_n) = \frac{K}{1 + a_1(j\omega_n) + b_1(j\omega_n)^2} \tag{4.24}$$

[3] Excepto para la aproximación de Chebyshev tipo II.

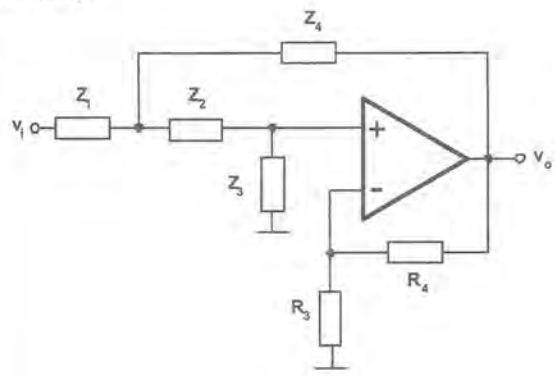
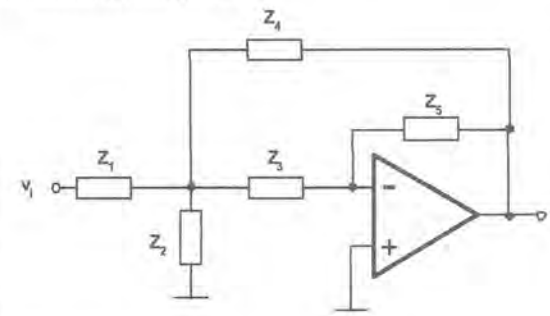
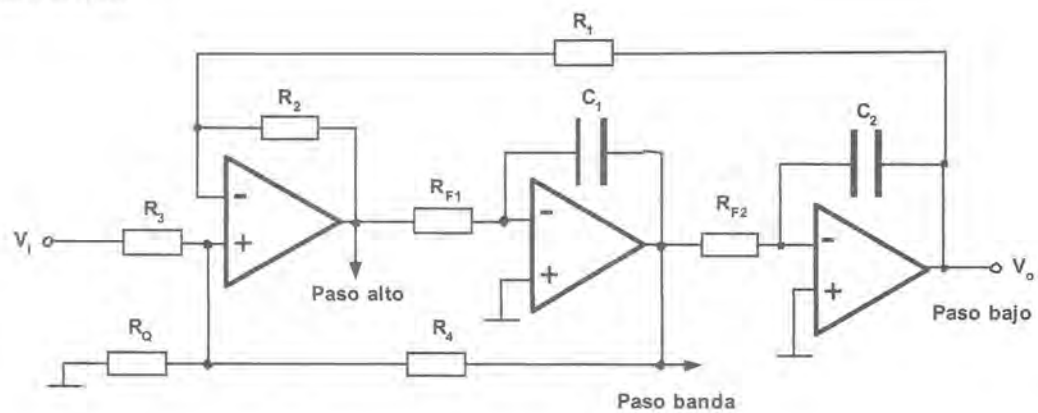
Estructuras para la realización de filtros activos	Características
Sallen-Key 	• No invierte la fase de la señal. • La menos dependiente de las características del operacional. • Es preferible cuando: – La exactitud en la ganancia es importante. – Filtros de ganancia unidad. – Los valores de Q son pequeños (Q<3). • En un filtro paso bajo: Z_1 y Z_2 resistencias; Z_3 y Z_4 condensadores. • En un filtro paso alto: Z_1 y Z_2 condensadores; Z_3 y Z_4 resistencias.
Realimentación múltiple (MFB) 	• Invierte la fase de la señal. • Poco sensible[4] a la variación de los componentes pasivos. • No es adecuada para filtros de Q alto debido a las limitaciones de ganancia del operacional en bucle abierto. • En un filtro paso bajo: Z_1, Z_3 y Z_4 resistencias; Z_2 y Z_5 condensadores. • En un filtro paso alto: Z_1, Z_3 y Z_4 condensadores; Z_2 y Z_5 resistencias.
Variables de estado  • Mayor precisión a costa de un mayor número de componentes. • Ajuste independiente de los parámetros del filtro. • Las funciones paso bajo, paso alto y paso banda se pueden obtener de forma simultánea. • Esta estructura se suele utilizar en forma de circuito integrado.	

Tabla 4.3. Principales estructuras físicas para la realización de filtros activos.

[4] La sensibilidad de un filtro es una medida de la vulnerabilidad de los parámetros del filtro (ganancia, frecuencia de corte y factor de calidad) a los cambios en los valores de los componentes. Este concepto se explicará en el apartado 4.9.

donde K es la ganancia del filtro, a_1, b_1 son los coeficientes que definen el tipo de aproximación matemática deseada (Anexo 2 en el CD adjunto) y ω_n la pulsación normalizada, definida como $\omega_n=\omega/\omega_c$. Para aclarar el método tomemos como ejemplo de diseño el filtro activo de primer orden de la figura 4.17.

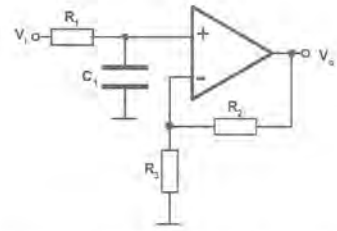


Figura 4.17. Filtro activo paso bajo no inversor de primer orden.

La función de transferencia de este circuito se puede poner de la forma:

$$H(j\omega)=\frac{1+(R_2/R_3)}{1+j\omega R_1C_1} \tag{4.25}$$

Sustituyendo ω por $\omega_c\cdot\omega_n$ se tiene:

$$H(j\omega_n)=\frac{1+(R_2/R_3)}{1+j\omega_c R_1C_1\omega_n} \tag{4.26}$$

Identificando los coeficientes de las expresiones (4.24) y (4.26), se obtiene:

$$K=1+(R_2/R_3)\ ;\ a_1=\omega_cR_1C_1\ ;\ b_1=0 \tag{4.27}$$

Como datos de partida en el diseño se suelen especificar la frecuencia de corte f_c, y la ganancia del filtro K. Se suele fijar el valor de C_1 ya que la gama de valores estándar de condensadores suele ser más reducida que la de resistencias. De esta forma, resolviendo las ecuaciones (4.27) respecto a R_1 y R_2, tenemos:

$$R_1=\frac{a_1}{2\pi f_cC_1}\ ;\ R_2=R_3(K-1) \tag{4.28}$$

El coeficiente a_1 se obtiene de las tablas del Anexo 2 en el CD adjunto, entrando con el tipo de aproximación que se desea tener y el orden del filtro, n. En el caso de un filtro de primer orden, a_1 siempre es la unidad, independientemente del tipo de aproximación.

EJERCICIO 4.4

Diseñe un filtro de primer orden de ganancia unidad y frecuencia de corte 1kHz.

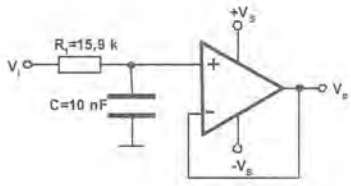
SOLUCIÓN:

Al tratarse de un filtro de primer orden el coeficiente a_1= 1. Seleccionando para C_1 el valor[5] de 10 nF, a partir de la expresión (4.28) se obtiene el valor de R_1:

[5] Para la selección de C_1 puede ser útil la siguiente regla práctica: $C_1(\mu F)=10/f_c$, donde f_c se expresa en Hz.

$$R_1 = \frac{a_1}{2\pi \cdot f_c \cdot C_1} = \frac{1}{2\pi \cdot 10^3\ \text{Hz} \cdot 10^{-8}\text{F}} = 15{,}9\ \text{k}\Omega$$

La figura 4.18 muestra los componentes del filtro, así como su respuesta frecuencial obtenida mediante simulación.



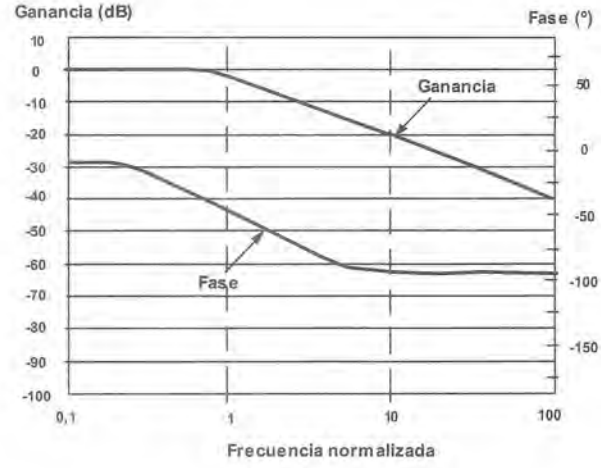


Figura 4.18

Estructura Sallen-Key

La figura 4.19 muestra la estructura general de un filtro paso bajo de segundo orden Sallen-Key.

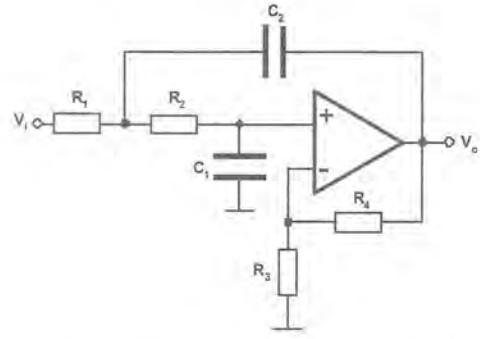


Figura 4.19. Estructura Sallen-Key de un filtro paso bajo de segundo orden.

Su función de transferencia, con respecto a la pulsación normalizada ω_n es:

$$H(j\omega_n) = \frac{K}{1 + \omega_c\left[C_1\left(R_1 + R_2\right) + \left(1 - K\right)R_1C_2\right] \cdot (j\omega_n) + \omega_c^2 R_1 R_2 C_1 C_2 \left(j\omega_n\right)^2} \qquad (4.29)$$

donde la ganancia es: $K = 1 + R_4/R_3$.

Es frecuente que la estructura Sallen-Key se utilice con ganancia unidad (K = 1), que es cuando proporciona una mayor exactitud. En este caso, la expresión (4.29) queda:

$$H(j\omega_n) = \frac{1}{1+\omega_c C_1(R_1+R_2)(j\omega_n)+\omega_c^2 R_1 R_2 C_1 C_2 (j\omega_n)^2} \tag{4.30}$$

Comparando los coeficientes de la ecuación (4.30) con los de la ecuación general (4.24) se obtiene el valor de los coeficientes a_1 y b_1:

$$a_1 = \omega_c C_1 (R_1+R_2); \qquad b_1 = \omega_c^2 R_1 R_2 C_1 C_2 \tag{4.31}$$

Fijando el valor de C_1 y de C_2, los valores de las resistencias R_1 y R_2 se obtienen por:

$$R_{1,2} = \frac{a_1 C_2 \mp \sqrt{a_1^2 C_2^2 - 4 b_1 C_1 C_2}}{4\pi \cdot f_c C_1 C_2} \tag{4.32}$$

Los valores dentro de la raíz han de ser reales, por lo que C_2 debe satisfacer la condición:

$$C_2 \geq C_1 \frac{4b_1}{a_1^2} \tag{4.33}$$

Un caso especial es tomar: $R_1 = R_2 = R$ y $C_1 = C_2 = C$. La función de transferencia (4.29) es ahora:

$$H(j\omega_n) = \frac{K}{1+\omega_c RC(3-K)\cdot(j\omega_n)+(\omega_c RC)^2 (j\omega_n)^2} \tag{4.34}$$

El filtro tiene ahora una ganancia K, ya que es físicamente imposible compatibilizar una ganancia unidad con que las resistencias y los condensadores sean iguales dos a dos, como se puede observar por las expresiones (4.32) y (4.33).

Comparando los coeficientes de la ecuación (4.34) con los de la ecuación general (4.24) se obtiene el valor de los coeficientes a_1 y b_1:

$$a_1 = \omega_c RC(3\text{-}K); \; b_1 = (\omega_c RC)^2 \tag{4.35}$$

Fijando C y despejando el valor de R se obtiene:

$$R = \frac{\sqrt{b_1}}{2\pi \cdot f_c C} \tag{4.36}$$

La relación entre la ganancia K, y el factor de calidad Q, está dada por [4.4]:

$$K = 3 - \frac{a_1}{\sqrt{b_1}} = 3 - \frac{1}{Q} \tag{4.37}$$

EJERCICIO 4.5

Diseñe un filtro paso bajo de segundo orden, de Butterworth con ganancia unidad y frecuencia de corte de 1 kHz.

SOLUCIÓN:

Del Anexo 2 del CD adjunto obtenemos los coeficientes de la aproximación de Butterworth: a_1 y b_1 para el caso de un filtro de segundo orden (n = 2), se obtienen así:

$$a_1 = 1{,}4142;\ b_1 = 1$$

Fijando $C_1 = 10$ nF, el valor de C_2 se obtiene por la expresión (4.33):

$$C_2 \geq C_1 \frac{4b_1}{a_1^2} = 10\ \text{nF} \cdot \frac{4 \cdot 1}{(1{,}4142)^2} = 20\ \text{nF}$$

Como $C_2 \geq 20$ nF tomamos para C_2 el valor estándar de 22 nF. Los valores de R_1 y de R_2 se obtienen de la expresión (4.32):

$$R_{1,2} = \frac{a_1 C_2 \mp \sqrt{a_1^2 C_2^2 - 4b_1 C_1 C_2}}{4\pi \cdot f_c C_1 C_2} = \frac{1{,}4142 \cdot 22 \cdot 10^{-9} \mp \sqrt{\left(1{,}4142 \cdot 22 \cdot 10^{-9}\right)^2 - 4 \cdot 1 \cdot 10^{-8} \cdot 22 \cdot 10^{-9}}}{4\pi \cdot 10^3 \cdot 10^{-8} \cdot 22 \cdot 10^{-9}} \Rightarrow$$

$$R_1 \approx 7\text{K86};\ R_2 \approx 14\text{K6}$$

El circuito final y los resultados de la simulación se muestran en la figura 4.20.

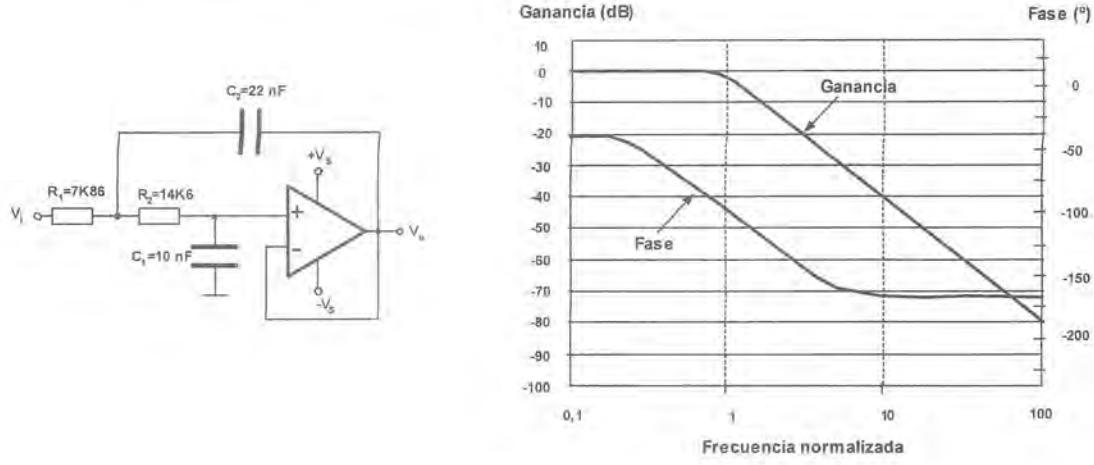


Figura 4.20

Estructura de realimentación múltiple (MFB)

La figura 4.21 muestra la estructura de realimentación múltiple de un filtro paso bajo. El operacional está en configuración inversora por lo que invierte la fase de la señal de entrada. La dependencia que tiene con los parámetros del operacional es mayor que en la realización Sallen-Key. Se suele emplear en el diseño de filtros que requieran valores de Q y de ganancia altos.

La función de transferencia es [4.4]:

$$H(j\omega_n) = \frac{-R_2/R_1}{1 + \omega_c C_1 \left(R_2 + R_3 + \frac{R_2 R_3}{R_1}\right) \cdot (j\omega_n) + \omega_c^2 C_1 C_2 R_2 R_3 \cdot (j\omega_n)^2} \qquad (4.38)$$

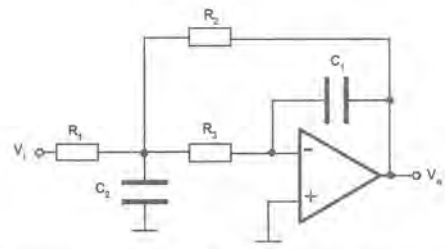


Figura 4.21. Estructura general de un filtro de segundo orden utilizando la estructura MFB.

Comparando los coeficientes de esta expresión con los de la expresión general (4.26) se obtienen las relaciones:

$$K = -\frac{R_2}{R_1}$$

$$a_1 = \omega_c C_1 \left(R_2 + R_3 + \frac{R_2 R_3}{R_1} \right) \qquad (4.39)$$

$$b_1 = \omega_c^2 C_1 C_2 R_2 R_3$$

Fijando C_1 y C_2 y resolviendo para R_1, R_2 y R_3 se obtienen las ecuaciones de diseño siguientes:

$$R_2 = \frac{a_1 C_2 - \sqrt{a_1^2 C_2^2 - 4 b_1 C_1 C_2 (1-K)}}{4\pi \ f_c C_1 C_2}$$

$$R_1 = -\frac{R_2}{k}$$

$$R_3 = \frac{b_1}{4\pi^2 f_c^2 C_1 C_2 R_2} \qquad (4.40)$$

$$C_2 \geq C_1 \frac{4 b_1 (1-K)}{a_1^2}$$

Filtros de orden superior

La realización de filtros de orden superior a dos se realiza conectando en cascada filtros de primer y de segundo orden. Si el filtro es de orden impar, la primera etapa será un filtro de primer orden y el resto de etapas filtros de segundo orden (figura 4.22a). En cambio, si el orden del filtro es par, todas las etapas parciales serán de segundo orden (figura 4.22b).

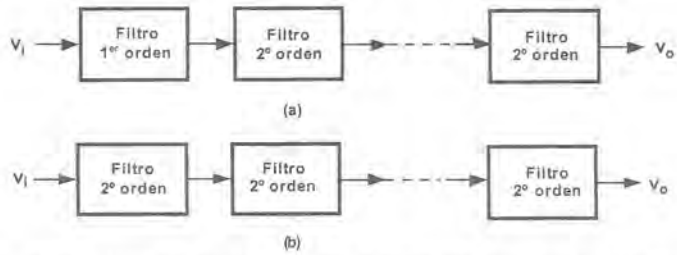


Figura 4.22. Formación de filtros de orden superior a partir de secciones de primer y segundo orden.

EJERCICIO 4.6

Diseñe un filtro paso bajo de Butterworth con ganancia unidad, de quinto orden y frecuencia de corte 1 kHz.

SOLUCIÓN:

La figura 4.23 muestra la conexión en cascada de un filtro de primer orden seguido de dos Sallen-Key de segundo orden para formar un filtro de quinto orden.

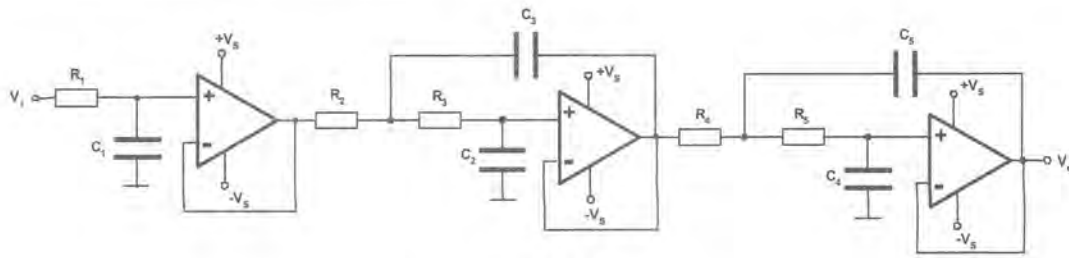


Figura 4.23

La tabla siguiente (extraída del Anexo 2 en el CD adjunto) nos muestra el valor que han de tomar los coeficientes a_i y b_i de la aproximación de Butterworth para cada etapa (i) del filtro de quinto orden (n = 5).

n	i	a_i	b_i	Q_i
5	1	1,0000	0,0000	–
	2	1,6180	1,0000	0,62
	3	0,6180	1,0000	1,62

Filtro 1: Los coeficientes de Butterworth de la primera etapa del filtro (i = 1) son, como ya sabemos: $a_1 = 1$; $b_1 = 0$.

Fijando $C_1 = 10$ nF, por la expresión (4.28) se obtiene: $R_1 = \dfrac{a_1}{2\pi f_c C_1} = \dfrac{1}{2\pi 10^3 \cdot 10^{-8}} = 15\text{K}9$

Filtro 2: Los coeficientes de Butterworth de la segunda etapa del filtro (i = 2): $a_2 = 1{,}6180$; $b_2 = 1$

Fijando $C_2 = 10$ nF, mediante la expresión (4.33) se puede seleccionar un valor para C_3:

$$C_3 \geq C_2 \frac{4b_2}{a_2^2} = 10\text{ nF} \cdot \frac{4 \cdot 1}{1{,}618^2} = 15{,}27\text{ nF}$$

Un valor posible es $C_3 = 18$ nF.

El valor de las resistencias R_2 y R_3 se obtiene de la expresión (4.32):

$$R_{2,3} = \frac{a_2 C_3 \mp \sqrt{a_2^2 C_3^2 - 4b_2 C_2 C_3}}{4\pi \cdot f_c C_2 C_3} = \frac{1{,}618 \cdot 18 \cdot 10^{-9} \mp \sqrt{\left(1{,}618 \cdot 18 \cdot 10^{-9}\right)^2 - 4 \cdot 10^{-8} \cdot 18 \cdot 10^{-9}}}{4\pi \cdot 10^3 \cdot 10^{-8} \cdot 18 \cdot 10^{-9}} \Rightarrow$$

$$R_2 = 7\text{K}8 \text{ y } R_3 = 17\text{K}7$$

Filtro 3: Los coeficientes de Butterworth de la tercera etapa del filtro (i = 3), son: $a_3 = 0{,}6180$; $b_3 = 1$.

Fijando $C_4 = 10$ nF:

$$C_5 \geq C_4 \frac{4b_3}{a_3^2} = 10\ \text{nF} \cdot \frac{4 \cdot 1}{0{,}618^2} = 104{,}7\ \text{nF}$$

Un posible valor para C_5 es 150 nF.

$$R_{4,5} = \frac{a_3 C_5 \mp \sqrt{a_3^2 C_5^2 - 4b_3 C_4 C_5}}{4\pi \cdot f_c C_4 C_5} = \frac{0{,}618 \cdot 150 \cdot 10^{-9} \mp \sqrt{\left(0{,}618 \cdot 150 \cdot 10^{-9}\right)^2 - 4 \cdot 10^{-8} \cdot 150 \cdot 10^{-9}}}{4\pi \cdot 10^3 \cdot 10^{-8} \cdot 150 \cdot 10^{-9}}$$

De donde: $R_4 = 2K2$ y $R_5 = 7K6$.

Con estos valores se ha realizado la simulación del filtro, obteniéndose los resultados que se muestran en la figura 4.24.

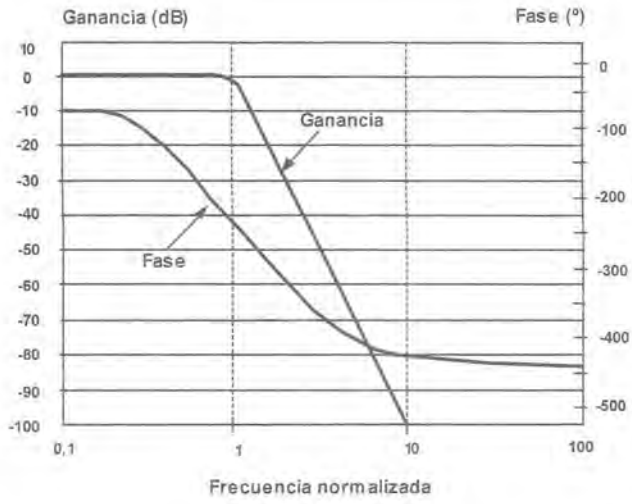


Figura 4.24

4.6.3 DISEÑO DE FILTROS PASO ALTO

La respuesta en frecuencia de un filtro paso alto se puede obtener sin más que reflejar la respuesta de un filtro paso bajo respecto a la frecuencia normalizada unidad, como se muestra en la figura 4.25. Esto da lugar a dos importantes consecuencias:

1. La función de transferencia de un filtro de segundo orden se obtiene a partir de la de un filtro paso bajo, expresión (4.26), reemplazando ω_n por $1/\omega_n$, es decir, la función de transferencia de un filtro paso alto es ahora:

$$H(j\omega_n) = \frac{K}{1 + \frac{a_1}{j\omega_n} + \frac{b_1}{(j\omega_n)^2}} \qquad (4.41)$$

2. La estructura física de un filtro paso alto se obtiene reemplazando en un filtro paso bajo las resistencias por condensadores y los condensadores por resistencias.

Para el diseño de los filtros paso alto, se seguirá el método expuesto para los filtros paso bajo, teniendo en cuenta las dos consideraciones anteriores. La tabla 4.4 muestra los circuitos y ecuaciones de diseño para los filtros paso alto de primer y segundo orden.

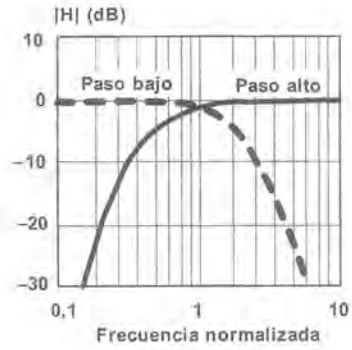


Figura 4.25. Respuesta frecuencial de un filtro paso alto a partir de un filtro paso bajo.

4.6.4 DISEÑO DE FILTROS PASO BANDA Y RECHAZO DE BANDA

Los filtros paso banda y rechazo de banda se especifican por tres parámetros característicos, que son:

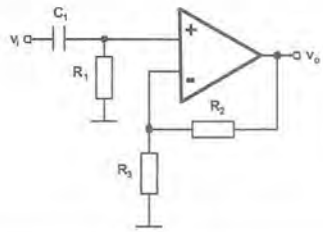
- Frecuencia central, f_0.
- Ganancia a la frecuencia central, K_0.
- Factor de calidad, Q.

El factor de calidad, como ya se comentó en la figura 4.5, es una indicación de la selectividad del circuito. Atendiendo al valor de Q, en ocasiones se habla de **filtros de banda ancha** ($Q < 10$) o de **filtros de banda estrecha** ($Q > 10$).

La función de transferencia de estos filtros se puede obtener también a partir de la de un filtro paso bajo, aunque ahora las transformaciones de frecuencia no sean tan evidentes. Por ello, la tabla 4.5 se ha limitado a proporcionar las ecuaciones de diseño y las características más significativas. En la referencia [4.4] se puede encontrar una demostración matemática de las mismas. En general, el circuito MFB es el más sencillo de todos, pero su empleo está restringido a valores de Q menores de 20. Su comportamiento está limitado por las características reales del operacional.

Diseño de filtros paso alto de primer orden

No inversor



Función de transferencia

$$H(j\omega_n)=\frac{1+(R_2/R_3)}{1+\frac{1}{\omega_c R_1 C_1}\cdot\frac{1}{j\omega_n}}$$

$$K=1+\frac{R_2}{R_3}\ ;\qquad a_1=\frac{1}{\omega_c R_1 C_1}$$

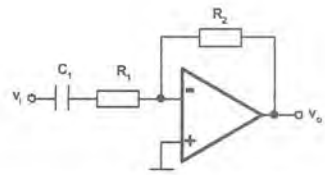
Ecuaciones de diseño

Datos: f_c, K
Fijar C_1

$$R_1=\frac{1}{2\pi\cdot f_c a_1 C_1}$$

$$R_2=R_3(K-1)$$

Inversor



Función de transferencia

$$H(j\omega_n)=\frac{-R_2/R_1}{1+\frac{1}{\omega_c R_1 C_1}\cdot\frac{1}{j\omega_n}}$$

$$K=-R_2/R_1$$

Ecuaciones de diseño

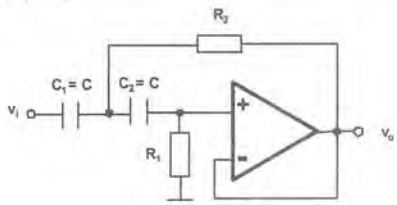
Datos: f_c, K
Fijar: C_1

$$R_1=\frac{1}{2\pi\cdot f_c a_1 C_1}$$

$$R_2=-R_1K$$

Diseño de filtros paso alto de segundo orden

Sallen-Key (K=1)



Función de transferencia

$$H(j\omega_n)=\frac{1}{1+\frac{2}{\omega_c RC}\cdot\frac{1}{j\omega_n}+\frac{1}{\omega_c^2 R_1 R_2 C^2}\cdot\frac{1}{(j\omega_n)^2}}$$

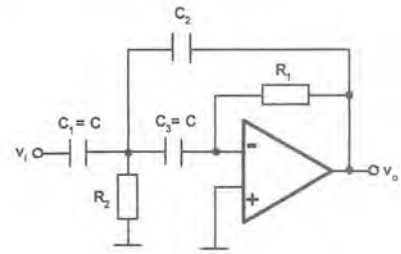
$$a_1=\frac{2}{\omega_c R_1 C};\ b_1=\frac{1}{\omega_c^2 R_1 R_2 C^2}$$

Ecuaciones de diseño

Datos: f_c. Seleccionar $C_1=C_2=C$.

$$R_1=\frac{2}{\omega_c C a_1}\ ;\ R_2=\frac{a_1}{2\omega_c C b_1}$$

MFB



Función de transferencia

$$H(j\omega_n)=\frac{-C/C_2}{1+\frac{2C+C_2}{\omega_c R_1 C C_2}\cdot\frac{1}{j\omega_n}+\frac{2C+C_2}{\omega_c R_1 C C_2}\cdot\frac{1}{(j\omega_n)^2}}$$

$$K=C/C_2\ ;\ a_1=\frac{2C+C_2}{\omega_c R_1 C C_2};\ b_1=\frac{2C+C_2}{\omega_c R_1 C C_2}$$

Ecuaciones de diseño

Datos: f_c, K. Seleccionar C y C_2.

$$R_1=\frac{1\text{-}2K}{\omega_c\cdot C\cdot a_1}\ ;\ R_2=\frac{a_1}{\omega_c\cdot b_1\cdot C_2(1-2K)}$$

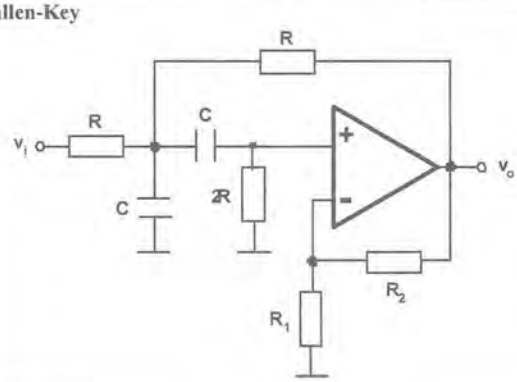
Tabla 4.4. Diseño de filtros paso alto de primer y de segundo orden.

Como ejemplo de filtro de rechazo de banda se ha seleccionado por su sencillez, únicamente el conocido **filtro en doble T**. Consta de un filtro pasivo en doble T (parte sombreada), a la que se le ha añadido un

operacional para poder obtener valores de Q más altos. Esto último se puede lograr de una manera más efectiva con los filtros de variables de estado que veremos a continuación.

Diseño de filtros paso banda de segundo orden

Sallen-Key



Relaciones de interés

$$\omega_0 = 1/RC \; ; \; K_0 = \frac{G}{3\text{-}G} \quad Q = \frac{1}{3\text{-}G}$$

siendo $G = 1 + \frac{R_2}{R_1}$

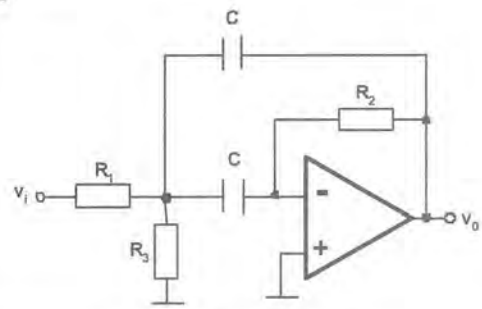
Ecuaciones de diseño

Datos: f_0, K_0, Q. Seleccionar C

$$R = \frac{1}{\omega_0 C} \text{ ó } R_2 = \frac{2Q-1}{Q} R_1$$

- Q puede variarse mediante G, sin modificar f_0.
- Q y K_0 no pueden ajustarse independientemente.

MFB



Relaciones de interés

$$\omega_0 = \frac{1}{C} \cdot \sqrt{\frac{R_1 + R_3}{R_1 R_2 R_3}} \; ; \; K_0 = R_2/2R_1$$

$$Q = \pi f_0 R_2 C \; ; \; B = 1/\pi R_2 C$$

Ecuaciones de diseño

Datos: f_0, K_0, Q. Seleccionar C

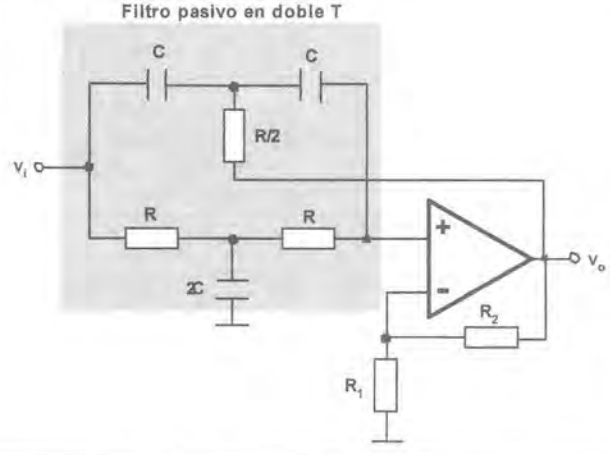
$$R_2 = \frac{Q}{\pi f_0 C} \; ; \; R_1 = \frac{R_2}{2K_0} \; ; \; R_3 = \frac{K_0 R_1}{2Q^2 + K_0}$$

- Permite ajustar Q, K_0 y f_0 independientemente.
- R_3 puede utilizarse para variar f_0 sin afectar a B o a K_0.

Diseño del filtros de rechazo de banda

Filtro en doble T

Filtro pasivo en doble T



Función de transferencia

$$\omega_0 = 1/RC \; ; \; K_0 = G \; ; \; Q = \frac{1}{2(2\text{-}G)}$$

siendo $G = 1 + \frac{R_2}{R_1}$

Ecuaciones de diseño

Datos: f_0, K_0

$$R = \frac{1}{2\pi f_0 C}$$

$$R_2 = (K_0 - 1)R_1 \text{ o } R_2 = R_1\left(1 - \frac{1}{2Q}\right)$$

- Q puede variarse mediante G, sin cambiar f_0.
- Q y K_0 no pueden ajustarse independientemente.

Tabla 4.5. Diseño de filtros paso banda y rechazo de banda de segundo orden.

4.7 Filtros activos de variables de estado

Las estructuras Sallen-Key y MFB empleadas hasta el momento para implementar filtros activos tienen tres importantes limitaciones:

- Son sensibles a las tolerancias de los componentes y de la temperatura.
- No permiten obtener valores de Q elevados.
- No permiten, en general, un ajuste independiente de la ganancia y del factor de calidad.

Los filtros de variables de estado permiten superar las limitaciones anteriores permitiendo alcanzar valores de Q más elevados (hasta 100) y ajustar todos parámetros del filtro de forma independiente a costa de un mayor número de componentes. La figura 4.26 muestra la estructura general de un filtro no inversor de variables de estado. Está compuesto por un amplificador sumador seguido de dos integradores. Se denominan también **filtros universales** ya que permiten obtener, salidas paso bajo, paso alto y paso banda. Si bien todas las salidas están disponibles simultáneamente, no lo están de forma optimizada, por lo que se suele utilizar bien como filtro paso alto, paso bajo o paso banda. Las aplicaciones típicas de este tipo de filtros se encuentran en equipos de comunicaciones, instrumentación médica, sistemas de adquisición de datos, etc.

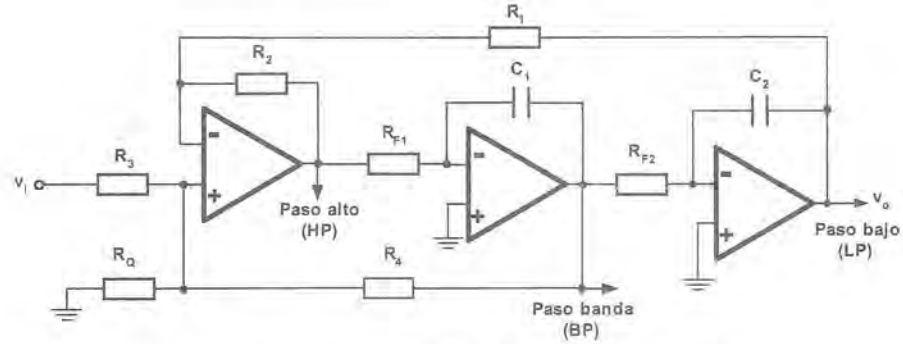


Figura 4.26. Filtro de variables de estado no inversor.

La versatilidad y buenas características generales de esta estructura han llevado a los fabricantes de circuitos integrados a integrar en una sola pastilla la mayor parte de los componentes del filtro, con lo que se consigue reducir la tolerancia de los componentes (±0,5%) y un ahorro de espacio en la placa de circuito impreso. La tabla 4.6 muestra las características de tres filtros comerciales, representativos de la estructura de variables de estado.

Referencia	Fabricante	Orden	f_0 máx.	V_{cc}
UAF42	Texas Instruments	2	100 kHz	±15 V
MAX275	Maxim	4	300 kHz	±5 V
LTC1562	Linear Technology	8	150 kHz	±5 V

Tabla 4.6. Filtros de variables de estado comerciales.

EJERCICIO 4.7

A partir del filtro de variables de estado UAF42 diseñe un filtro de ganancia unidad que elimine una interferencia de 50 Hz que se encuentra superpuesta a una determinada señal objeto de lectura. El ancho de banda de rechazo del filtro ha de ser como máximo, de 2 Hz.

SOLUCIÓN[6]:

Se trata de diseñar un filtro que rechace la frecuencia de 50 Hz y deje pasar el resto de frecuencias. El UAF42 dispone además de los tres operacionales típicos de la estructura de variables de estado, un cuarto operacional auxiliar que se puede utilizar para implementar un filtro de rechazo de banda como se muestra en la figura 4.27. Para ello se suman las salidas paso alto y paso bajo, de forma tal que a la frecuencia f_0 ambas salidas son de igual magnitud pero desfasadas 180°, por lo que su suma será cero, rechazándose la frecuencia de 50 Hz.

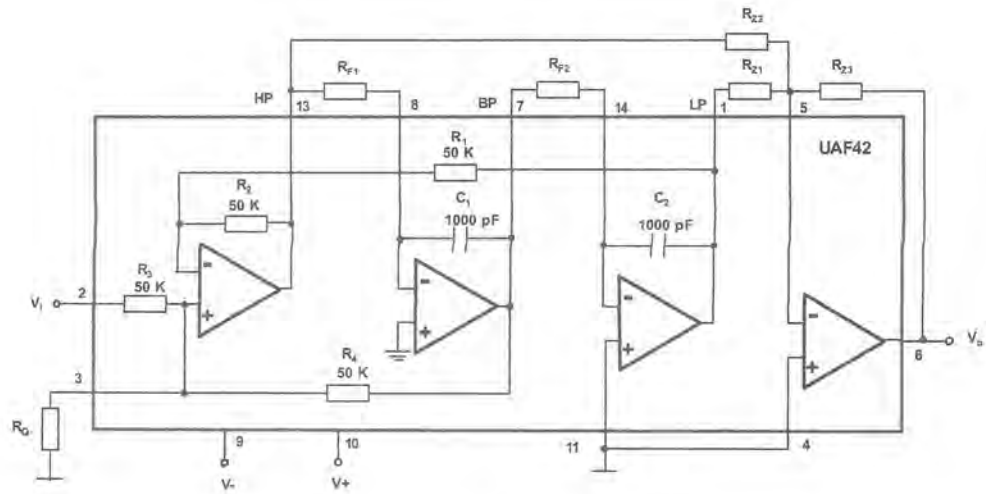


Figura 4.27

Las ecuaciones de diseño se simplifican notablemente tomando: $R_{F1} = R_{F2} = R$ y $C_1 = C_2 = C$ [4.7]:

$$\omega_0 = \frac{1}{R \cdot C};\ Q = \frac{1}{2}\left[1 + \left(\frac{50K + R_Q}{R_Q}\right)\right] \tag{4.42}$$

Despejando R_Q se obtiene:

$$R_Q = \frac{25\ k\Omega}{Q-1} \tag{4.43}$$

Tomando C = 1 nF, de la expresión (4.43), para f_0 = 50 Hz resulta $R = R_{F1} = R_{F2} = 3{,}18\ M\Omega$. Como el ancho de banda ha de ser igual o inferior a 2 Hz, el factor de calidad, $Q = f_0/B$, ha de ser igual o superior a

[6] Se puede emplear el programa FILTER42 de Texas Instruments, disponible en su página Web: http://www.ti.com. Este programa permite seleccionar los valores de los componentes para los tres tipos de aproximaciones matemáticas vistas.

25. Para Q = 25 resulta $R_Q \cong 1$ kΩ. Las resistencias de las entradas al sumador (R_{Z1} y R_{Z2}) y la resistencia de realimentación (R_{Z3}) han de cumplir las relaciones [4.8]:

$$Q = \frac{R_{Z3}}{R_{Z1}} = \frac{R_{Z3}}{R_{Z2}} \tag{4.44}$$

Para Q = 25 y $R_{Z1} = R_{Z2} = 2$ kΩ, resulta $R_{Z3} = 50$ kΩ. La figura 4.28 muestra los resultados de la simulación.

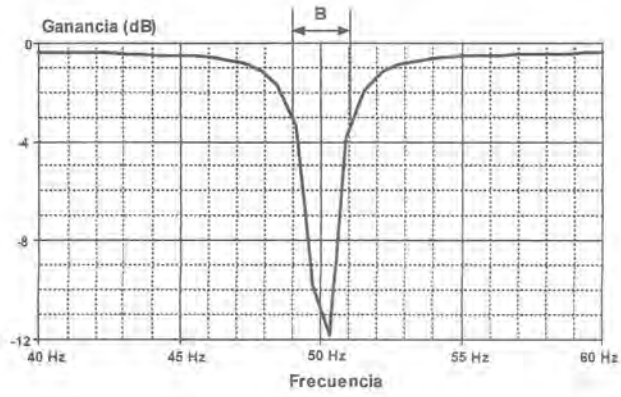


Figura 4.28

4.8 Filtros de capacidad conmutada

La idea básica de los filtros de capacidades conmutadas consiste en reemplazar las resistencias del filtro por interruptores MOS y condensadores integrados en el propio chip de silicio. De esta manera se consiguen dos objetivos: por un lado, integrar resistencias de alto valor (con disipación de potencia insignificante) y por otro, poder ajustar el valor de dichas resistencias mediante la variación de la frecuencia de conmutación de los interruptores. El valor de la frecuencia de corte estará determinado y será proporcional a la frecuencia de un reloj externo. A diferencia de los filtros convencionales la señal de entrada es muestreada y procesada en un tiempo discreto.

Para explicar el principio de funcionamiento de este tipo de filtros, la figura 4.29 muestra un integrador estándar basado en un operacional. Cuando se aplica una tensión de entrada positiva, la salida del integrador decrecerá a una velocidad determinada por la tensión de entrada y la constante de tiempo RC. La rampa tiene una pendiente de valor:

$$\frac{\Delta v}{\Delta t} = -\frac{v_i}{RC}$$

La precisión de la constante de tiempo y, por tanto, la del integrador (filtro paso bajo de primer orden) depende de la precisión absoluta de dos componentes discretos completamente diferentes. El objetivo no es otro que hacer que la constante de tiempo sea lo menos dependiente de las características de los componentes. Veamos cómo se logra esto con la técnica de las capacidades conmutadas.

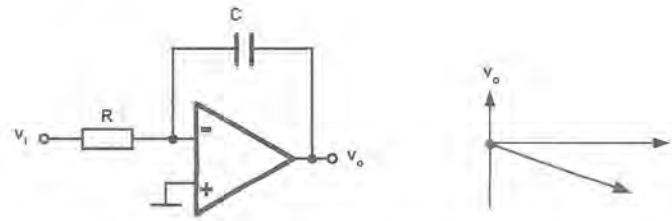


Figura 4.29. Integrador RC mediante un operacional.

Para ello la figura 4.30 muestra un integrador inversor de capacidad conmutada, donde la R del integrador de la figura 4.29 ha sido reemplazada por un condensador C_1 y dos interruptores analógicos, S_1 y S_2. Los interruptores S_1 y S_2 se cierran y abren de forma alternativa por medio de un reloj. Cuando S_1 está cerrado y S_2 abierto, C_1 se carga al valor de v_i. Al final de medio periodo de reloj, la carga almacenada en C_1 es v_iC_1. Cuando el reloj cambia de estado, S_1 se abre y S_2 se cierra. Durante este semiperiodo del reloj toda la carga almacenada en C_1 se transfiere al condensador de realimentación C_2 y la corriente que circula por él vale:

$$I = \frac{\Delta Q}{\Delta t} = \frac{v_i C_1}{T} = v_i C_1 f_{CLK}$$

donde T es el periodo del reloj.

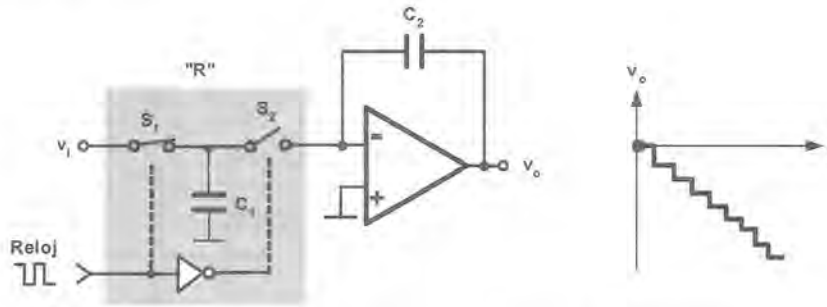


Figura 4.30. Integrador inversor de capacidad conmutada.

La resistencia efectiva vista desde v_i a la entrada (–) del operacional es:

$$"R" = \frac{v_i}{I} = \frac{1}{C_1 \cdot f_{CLK}} \tag{4.45}$$

Esto significa que S_1, S_2 y C_1 actúan de la misma forma que la resistencia R de la figura 4.29. La constante tiempo vale ahora:

$$"R"C_2 = \frac{C_2}{C_1 \cdot f_{CLK}} \tag{4.46}$$

que depende sólo de la relación entre los valores de dos condensadores. Es posible ajustar su valor fácilmente integrando ambos condensadores en el mismo chip.

Así, para un filtro con estructura de variables de estado implementado con tecnología de capacidad conmutada, la frecuencia central está dada por la expresión 4.42:

$$f_0 = \frac{1}{2\pi \text{"R"} C_2} = \frac{f_{CLK} \cdot C_1}{2\pi \cdot C_2} \qquad (4.47)$$

La frecuencia central depende ahora de la frecuencia del reloj y de la relación de los valores de los condensadores y no de sus valores absolutos. La estabilidad de f_0 es tan buena como lo sea la del reloj, por lo que si se emplean osciladores de cristal la estabilidad será muy alta. Otra conclusión importante que se desprende de la expresión (4.47) es que podemos ajustar la frecuencia crítica controlando la frecuencia del reloj (filtros sintonizables).

Los interruptores se implementan con tecnología MOS como se muestra en la figura 4.31. Una resistencia de 10 MΩ, se puede integrar mediante un condensador de 1 pF conmutando a 100 kHz. El área de silicio que se precisa es aproximadamente 0,01 mm^2. En cambio, la resistencia equivalente implementada mediante técnicas de difusión requeriría un área de silicio 100 veces mayor.

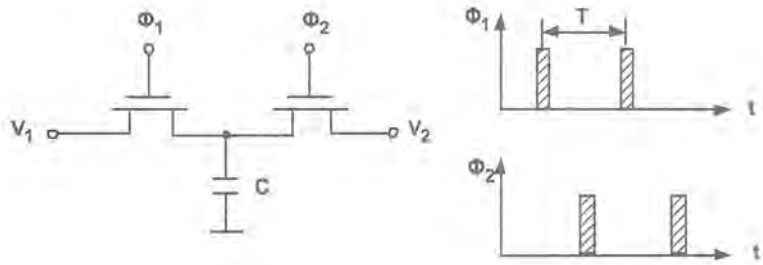


Figura 4.31. Implementación MOS de un condensador conmutado.

Los filtros de capacidades conmutadas no están exentos de problemas. Las capacidades conmutadas y los operacionales introducen ruido aleatorio y corrientes de dispersión que producen errores de *offset*. Por otro lado, una parte de la señal de reloj (del orden de 10 mVp-p) se transmite a la salida del filtro y aunque su frecuencia es mucho mayor que la frecuencia de corte del filtro puede ocasionar problemas en aplicaciones de bajo ruido. Al trabajar con señales muestreadas se ha de cumplir el teorema del muestreo, que obliga a que la frecuencia de muestreo, f_{CLK}, sea, como mínimo, mayor que el doble de la máxima frecuencia de la señal de entrada. De no cumplirse este criterio, la salida del filtro aparecerá distorsionada. Esta distorsión es debida a un fenómeno conocido como *aliasing*, el cual se estudiará en el capítulo 22.

Comercialmente se dispone de filtros de capacidades conmutadas de función fija y de tipo universal. Los primeros, tanto la función (paso bajo, paso alto, etc.) como la aproximación matemática, son fijos. En cambio, en los filtros de tipo universal, la función de salida es configurable en función de resistencias externas, estando basados en una estructura de variables de estado. Como ejemplo de filtro de función fija, la figura 4.32 muestra el patillaje del LMF40 de National Semiconductor. Se trata de un filtro paso bajo, de Butterworth, de cuarto orden. La frecuencia de corte se puede establecer en el margen de 0,1 Hz a 40 kHz. La relación con la frecuencia de reloj está dada por:

$$\frac{f_{CLK}}{f_c} = 50 \text{ (para el LMF40-50) o bien } \frac{f_{CLK}}{f_c} = 100 \text{ (para el LMF40-100)}$$

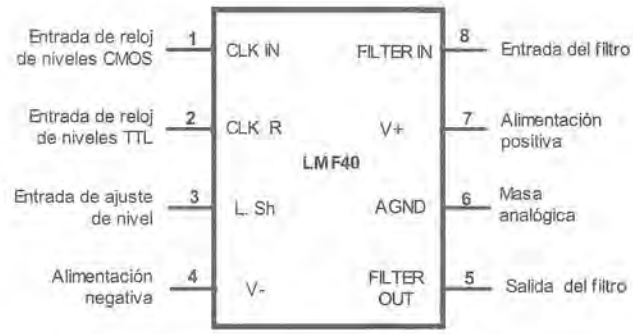


Figura 4.32. Patillaje del filtro LMF40 de National Semiconductor.

Para aplicaciones de bajo coste la señal de reloj puede generarse a partir de un circuito *Schmitt-Trigger* interno y una red RC externa, como se muestra en la figura 4.33.

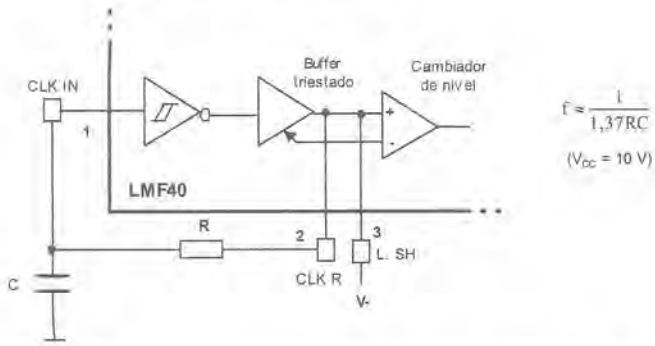


Figura 4.33. Circuito oscilador RC *Schmitt-Trigger* del LMF40.

EJERCICIO 4.8

A partir del circuito integrado LMF40 diseñe un filtro paso bajo de Butterworth que cumpla las siguientes especificaciones: $A_{min} = 20$ *dB,* $A_{máx} = 2{,}0$ *dB,* $f_p = 1$ *kHz y* $f_s = 2$ *kHz.*

SOLUCIÓN:

Las especificaciones del ejemplo se han representado en la figura 4.34:

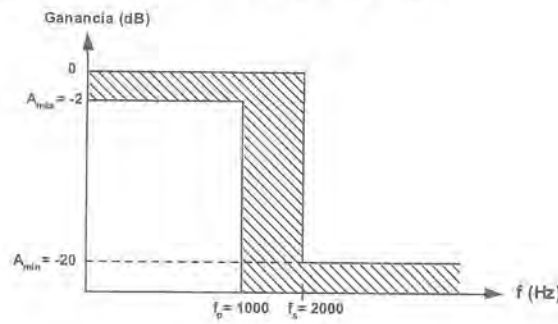


Figura 4.34

Veamos cuál es el orden del filtro de Butterworth que puede cumplir las especificaciones anteriores. Este orden se calcula por las expresiones (4.10) y (4.12):

$$\epsilon = \sqrt{10^{A_{max}/10} - 1} = \sqrt{10^{2/10} - 1} = 0,7647$$

$$n = \frac{\log\left[\left(\frac{1}{A_{min}^2}\right) - 1\right] - 2\log \epsilon}{2\log\left(\frac{f_s}{f_p}\right)} = \frac{\log\left[\left(\frac{1}{0,1^2}\right) - 1\right] - 2\log 0,7647}{2\log(2)} = 3,7$$

Como n sólo puede tomar valores enteros, el orden del filtro es n = 4, por lo que es posible implementar el filtro con un solo circuito LMF40. La frecuencia de corte, f_c, se obtiene a partir de la expresión (4.7):

$$|H| = \frac{1}{\sqrt{1 + \epsilon^2\left(\frac{f_c}{f_p}\right)^{2n}}} = 0,7071 \quad (-3\,dB)$$

donde: n = 4, ϵ = 0,764, f_p = 1 kHz. Despejando la frecuencia de corte se obtiene f_c = 1,068 kHz. La frecuencia del reloj es igual a: $f_{CLK} = 50 \cdot f_c = 50 \cdot 1,068 = 53,4$ kHz (modelo LMF40-50) o bien: $f_{CLK} = 100 \cdot f_c = 100 \cdot 1,068 = 106,8$ kHz. (Modelo LMF40-100). La figura 4.35 muestra el diagrama del LMF40 empleando una señal de reloj de niveles TTL.

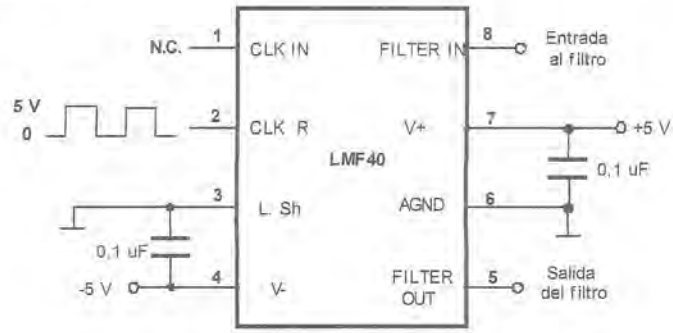


Figura 4.35

4.9 Selección de componentes

Hasta este momento las características de todos los componentes electrónicos utilizados en los filtros se han considerado ideales; sin embargo, la tolerancia de los condensadores y de las resistencias seleccionadas en el diseño puede afectar fuertemente a los parámetros del filtro, principalmente al factor de calidad, Q y a la frecuencia de corte, f_c. La **sensibilidad** de un valor Q, que cambia con respecto al valor de la resistencia R, se define como:

$$S_R^Q \approx \frac{\Delta Q/Q}{\Delta R/R} \tag{4.48}$$

De forma similar se definen las sensibilidades siguientes: S_C^Q, S_R^f, S_C^f. Lógicamente interesará que la sensibilidad sea lo menor posible. Para las estructuras Sallen-Key y MFB de segundo orden, las sensibilidades anteriores son del orden de ±0,5%/%. Esto significa que si, por ejemplo, el valor de la resistencia varía en un 1%, según la expresión (4.48) tenemos:

$$\frac{\Delta Q}{Q} = 0{,}5 \cdot \frac{\Delta R}{R} = 0{,}5 \cdot 1 = 0{,}5\%$$

Aunque estos valores no son grandes, en el caso de filtros de orden superior, las diferencias de Q y f_c, de cada filtro parcial pueden hacer que la respuesta del filtro completo se vea fuertemente alterada.

En la figura 4.36a se muestran 10 simulaciones realizadas con el programa PSpice[7] [4.8] de la respuesta del filtro *notch* del ejemplo 4.7, tomando una tolerancia para los componentes pasivos del 1%. Vemos cómo, a pesar de los valores de tolerancia reducidos, se producen desviaciones de sintonización con respecto a la frecuencia central de 50 Hz. Estas desviaciones se ven más acusadas cuando la tolerancia de los condensadores es del 5% (figura 4.36b). Este tipo de análisis estadístico se conoce como **análisis de Montecarlo** y es empleado frecuentemente en la producción de circuitos para predecir la respuesta del circuito ante variaciones en el valor de los componentes.

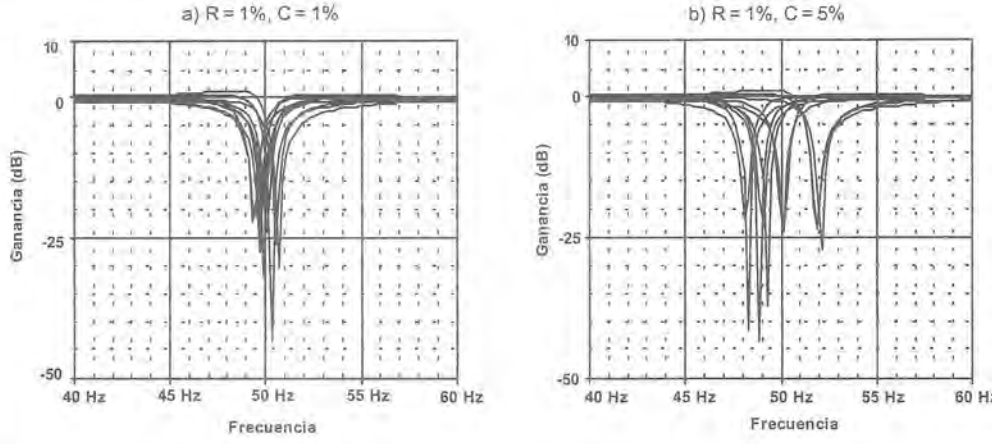


Figura 4.36. Respuesta de un filtro *notch* de 50 Hz teniendo en cuenta la tolerancia de los componentes.

Selección de resistencias

Cuando se trabaja con operacionales, los valores de las resistencias deberán estar en la banda de 1 a 100 kΩ. El límite inferior evita forzar a que el operacional maneje una corriente excesiva en su salida, lo cual es particularmente importante en el caso de operacionales alimentados con tensión única. El límite superior de 100 kΩ garantiza que el ruido térmico generado no sea excesivo, dado que se incrementa con la raíz cuadrada del valor de la resistencia.

Por otro lado, es importante considerar el comportamiento de las resistencias en alta frecuencia, para ello la figura 4.37 muestra su circuito equivalente en alterna. Hay que tener en cuenta que aunque el circuito no

[7] En PSpice una tolerancia del 1% en el valor de una resistencia se especifica en el modelo de la resistencia mediante la expresión: *.model Rmodel RES R=1 DEV=1%*. De la misma forma una tolerancia del 5% en el condensador se especifica en el modelo del condensador como *.model Cmodel CAP C=1 DEV=5%*. Los nombres *Rmodel* y *Cmodel* pueden cambiarse.

maneje señales de alta frecuencia, estas pueden ser radiadas o conducidas al circuito. A bajas frecuencias, tanto el efecto inductivo (debido la longitud de los terminales) como el capacitivo (debido a la capacidad parásita entre terminales separados una cierta distancia separados por un dieléctrico) no son apreciables y pueden despreciarse. Sin embargo, a medida que la frecuencia aumenta la capacidad parásita, C_p, aparece en paralelo, aumentando con el valor de la resistencia. Como consecuencia, la impedancia equivalente de la resistencia se ve reducida. En estas condiciones el margen de valores de resistencias aconsejable es aún más reducido, entre 1 y 10 kΩ. Una buena opción son las resistencias de **película metálica** (*metal film*). Se caracterizan por tener un amplio margen de valores comerciales, tolerancias de precisión (±0,1%), coeficiente de temperatura excelente (<1 a 100 ppm/°C), así como un coste moderado. Más aconsejables aún son las resistencias de montaje superficial de película gruesa, donde la inductancia se reduce al mínimo.

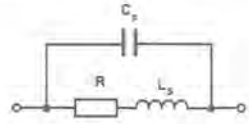


Figura 4.37. Circuito equivalente de una resistencia en alta frecuencia.

Selección de condensadores

En el caso de la selección de los condensadores la situación se complica un poco más debido a la gran diversidad de tipos existentes. Dado que no existe tanta variedad de valores de condensadores como de resistencias, se recomienda fijar primero el valor del condensador. En el caso de no disponer de condensadores de baja tolerancia se puede recurrir a medir el valor del condensador y de acuerdo a él calcular el valor de la resistencia necesaria. Los valores adecuados para aplicaciones de filtrado suelen estar en el margen de 1 nF a 1 μF.

Las características más importantes a considerar en la elección del tipo de condensador son: baja inductancia propia, bajo coeficiente de pérdidas, tolerancia máxima de ±10% y alta resistencia de aislamiento. En general se recomiendan utilizar condensadores de alguno de los siguientes tipos:

- Cerámicos multicapa (con dieléctricos COG, X7R).
- Película plástica (poliestireno, polipropileno o poliéster).
- Mica.

Se deben evitar condensadores electrolíticos debido a su alta tolerancia y a la alta resistencia serie equivalente (ESR) que presentan. Sólo en el caso de ser necesarios valores de capacidad altos se emplearan condensadores electrolíticos de tántalo[8]. Una forma de reducir componentes parásitos es emplear condensadores para montaje superficial. Al reducirse el tamaño de los terminales, la inductancia de los mismos se reduce drásticamente. Además, su reducido tamaño permite una óptima colocación. Una desventaja es que no todos los tipos de condensadores están disponibles en montaje superficial. Los condensadores cerámicos son muy populares en montaje superficial, y de estos los multicapa COG, con un coeficiente de temperatura prácticamente nulo, son una buena selección.

Efecto de la temperatura

Otro problema importante es la influencia de la temperatura en el valor de las resistencias y de la capacidad de los condensadores. En ocasiones, los equipos han de poder trabajar correctamente en un margen de temperaturas amplio. Es aconsejable que su coeficiente de temperatura sea cero o lo menor posible. Para poner de manifiesto esto se ha recurrido a realizar la simulación del filtro paso bajo de segundo orden del ejercicio 4.5. Se trata de un filtro de Butterworth con ganancia unidad y frecuencia de corte de 1 kHz. Los resultados

[8] Los condensadores de tántalo están polarizados, soportando tensión sólo en un sentido. Para evitar este problema suelen emplearse dos condensadores en serie con polaridades opuestas.

se muestran en la figura 4.38. La dependencia de la temperatura de las resistencias y de los condensadores se modela añadiendo el parámetro TC = 0,01 en la declaración del modelo. La frecuencia de corte varía desde aproximadamente 5 Hz a –15°C hasta los 700 Hz a 45°C. Los resultados obtenidos corresponden a resistencias y condensadores con coeficiente de temperatura positivo. En este caso dichos valores aumentarán con la temperatura, en tanto que la frecuencia de corte disminuirá. Recuerde que la frecuencia de corte viene dada por la expresión:

$$f_c = 1/2\pi\sqrt{R_1R_2C_1C_2}$$

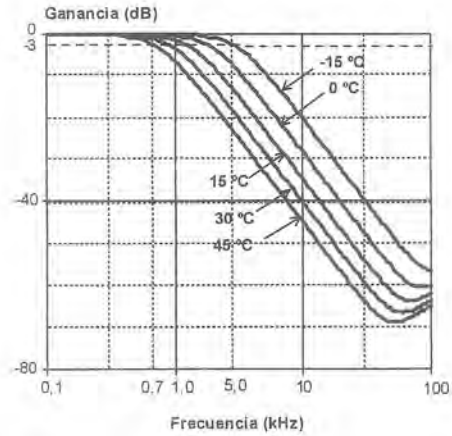


Figura 4.38. Respuesta de un filtro de Butterworth de 2° orden a diferentes valores de temperatura.

4.10 Programas de ayuda al diseño de filtros

El diseño de un filtro se facilita sustancialmente empleando herramientas software de diseño de filtros. La tabla 4.7 recoge algunos de los programas de distribución gratuita proporcionados por los fabricantes de circuitos integrados a través de su página Web.

Programa	Fabricante	Características	Limitaciones
FilterLab	Microchip (www.microchip.com)	• Diseño de filtros hasta de 8° orden. • Diseño de filtros *anti-aliasing*. • Genera salida para Spice.	• Sólo filtros activos paso bajo Sallen-Key o MFB. • No permite respuesta temporal.
FilterCAD	Linear Technology (www.linear.com)	• Todo tipo de filtros, respuestas y estructuras. • Respuesta frecuencial y temporal.	• Sólo para circuitos integrados de Linear. • No genera salida para Spice.
MAX274	Maxim (www.maxim-ic.com)	• Todas las aproximaciones vistas. • Filtro de variables de estado universal.	• Sólo para DOS. • Sólo para el MAX274 y 275.
Filter2	Texas Instruments (www.ti.com)	• Permite las 4 funciones y las tres aproximaciones. • Estructuras Sallen-Key y MFB.	• Sólo para DOS.
Filter42	Texas Instruments (www.ti.com)	• Filtro Universal UAF42.	• Sólo para DOS.

Tabla 4.7. Programas de ayuda al diseño de filtros analógicos.

4.11 Comparación de tecnologías de filtros

Una vez presentadas por separado las características más importantes de las diferentes tecnologías de filtros (pasivos, activos y de capacidad conmutada), vamos a realizar un estudio comparativo de las mismas atendiendo a diferentes criterios [4.5]:

- **Exactitud.** Los filtros de capacidad conmutada son los que tienen, en general, una mayor exactitud. Así, la frecuencia central se puede obtener con una exactitud alrededor del 0,2%. Para alcanzar este porcentaje con los filtros pasivos o activos se tendrían que emplear resistencias y condensadores de muy alta precisión lo cual encarecería su coste.

- **Coste.** Si el criterio es el coste, la elección de la tecnología más adecuada no está muy clara. En general, podemos decir que si sólo se requiere un polo, una red RC puede ser la solución ideal. Para diseños más complejos los filtros de capacidad conmutadas pueden ser económicos y ocupar poco espacio en la placa de circuito impreso. Si se requiere una buena exactitud, los componentes pasivos, especialmente los condensadores, empleados en las técnicas discretas pueden ser bastante costosos. Cuando la velocidad y la exactitud no son aspectos demasiado importantes, pueden resultar más económicos los filtros activos convencionales.

- **Ruido.** Los filtros pasivos generan poco ruido (sólo ruido de origen térmico debido a las resistencias). A continuación, están los filtros activos convencionales y después los de capacidades conmutadas. En estos últimos, a pesar de que las resistencias de entrada están construidas mediante interruptores y condensadores, producen ruido térmico como si se tratasen de resistencias "reales". Además del ruido térmico, en los filtros de capacidades conmutadas, una parte de la señal de reloj (del orden de 10 mVp-p) se transmite a la salida del filtro.

- **Tensión de *offset*.** Los filtros pasivos no tienen tensión de *offset*. En los filtros activos la tensión de *offset* es función de la tensión de *offset* de los operacionales y de la ganancia en continua de las etapas de filtrado. La elección cuidadosa del operacional ayudará en este sentido. Los filtros de capacidad conmutada son los que tienen las tensiones de *offset* mayores (desde pocos mV hasta 100 mV) por lo que no son adecuados para aquellas aplicaciones donde se requiera una buena precisión en continua, al menos si no se utiliza circuitería externa para corregir este *offset*.

- **Margen de frecuencia.** Los filtros de capacidad conmutada pueden cubrir sin problemas un margen de frecuencias entre 0,1 y 100 kHz aproximadamente. Para los filtros pasivos y los activos trabajar a muy bajas frecuencias supone el empleo de componentes reactivos grandes y, en general, costosos. A frecuencias altas, por encima de los 100 kHz, los operacionales han de ser rápidos y, por tanto, más caros.

- **Ajuste.** En los filtros activos y pasivos, un cambio en la frecuencia de corte afecta al valor de algunos de los componentes. Por contra, en los filtros de capacidad conmutada, la frecuencia central del filtro (o frecuencia de corte) es proporcional a la frecuencia del reloj, pudiendo seleccionarse entre 5 y 6 décadas, sin alterar la circuitería externa. Esto puede ser importante en aplicaciones donde se requieran múltiples frecuencias de corte.

- **Número de componentes/Área de tarjeta.** Desde este punto de vista los filtros de capacidad conmutada son los ganadores. Los filtros de capacidad conmutada de función fija (dedicados) no requieren componentes externos (excepto el reloj) y los universales (programables) requieren entre 2 y 4 resistencias para cada función de segundo orden, ocupando menos espacio que las otras alternativas. Los filtros pasivos necesitan un condensador por polo y los activos requieren al menos un operacional, dos *resistencias y dos condensadores para cada filtro de segundo orden.*

- **Esfuerzo de diseño.** En general, los filtros de capacidad conmutada son los más fáciles de diseñar. En algunos casos, como hemos visto con el LMF40 de National, sólo se requiere la señal de reloj. Es muy recomendable el empleo de herramientas software, las cuales, además de simplificar el diseño permiten visualizar la respuesta frecuencial del filtro.

4.12 Aplicaciones de los filtros

Los filtros analógicos encuentran aplicación en numerosos campos de la electrónica. En este punto se tratarán sólo algunas de sus aplicaciones más representativas, así como el tipo de filtro más adecuado en cada caso [4.5].

- **Aplicaciones de eliminación de ruido:** filtros *notch* para eliminar la frecuencia de línea (por ejemplo, en instrumentación biomédica), filtros paso bajo para la eliminación de ruido en aplicaciones de instrumentación general, filtros *anti-aliasing*[9]; por ejemplo, en los sistemas de adquisición de datos.

  Todas estas aplicaciones pueden ser implementadas tanto mediante filtros de capacidad conmutada como con filtros activos convencionales. Cuando se requiere tener buena precisión en continua se suele acudir a filtros activos, construidos con amplificadores operacionales de precisión.

- **Aplicaciones de detección de tono:** comunicaciones, fax, módem, instrumentación biomédica, instrumentación acústica, etc.

  En este tipo de aplicaciones los filtros de capacidad conmutada son siempre la mejor elección en virtud de la mayor exactitud en el valor de la frecuencia central. Además, ocupan poco espacio en la placa de circuito impreso.

- **Aplicaciones de filtrado con frecuencia ajustable:** análisis de espectros, filtros multifunción, etc.

  En este tipo de aplicaciones, donde se requieran múltiples frecuencias centrales, los filtros más adecuados son los de capacidad conmutada. Las frecuencias centrales (o de corte) son controladas por la frecuencia del reloj, pudiendo llegar a cubrir un margen de 5 décadas. En los filtros de tiempo continuo, el ajuste de la frecuencia de corte es mucho más difícil y requiere interruptores analógicos para conmutar entre componentes resistivos.

- **Aplicaciones de procesamiento de señales de audio:** controles de tono, filtros paso alto, redes *crossover* activas, etc.

  Los filtros de capacidades conmutadas son, en general, demasiado ruidosos para aplicaciones de audio de alta fidelidad. Los filtros de audio normalmente requieren manejar tres décadas de frecuencias, pudiendo tener problemas de *aliasing*. En general, los filtros de tiempo continuo son más adecuados para aplicaciones de audio. Algunos sistemas de comunicación tienen anchos de banda y relaciones señal/ruido que son compatibles con filtros de capacidad conmutada, con la ventaja del ajuste y del menor tamaño.

EJERCICIO 4.9

Se pretende realizar el diseño de un circuito de precisión para la medida de temperatura. Razone el tipo de filtro que se podría implementar para este tipo de aplicación.

[9] Los filtros *antialising* se estudiarán en el capítulo 22. Su función es doble: por un lado, limitar el ancho de banda de las señales de interés al estrictamente necesario y, por otro lado, limitar el espectro del ruido y otras interferencias que a menudo alteran la señal deseada.

SOLUCIÓN:

La temperatura es una variable física en la que los cambios se producen de forma muy lenta. Por otro lado, la constante térmica de los elementos sensores impide que su respuesta sea rápida. Podemos decir que, en general, los circuitos de medida de temperatura operan prácticamente en continua. La estrategia de diseño del filtro ha de ser tal que su frecuencia de corte sea baja y la región de transición suficientemente estrecha. La frecuencia de corte, f_c, se puede se situar en 1 Hz, lo cual permite eliminar el ruido generado por los componentes electrónicos. Además, a esta frecuencia los valores de los condensadores son todavía razonables. La frecuencia extrema de la banda de transición, f_s, deberá ser 50 Hz o inferior, a fin de eliminar el ruido de línea de 50 Hz.

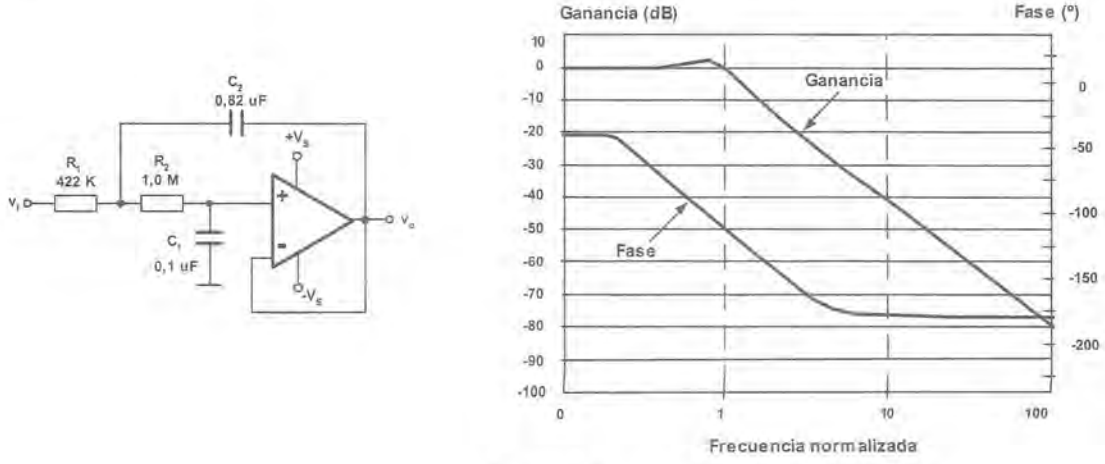


Figura 4.39

De las tres aproximaciones estudiadas, la aproximación de Chebyshev sabemos que es la que presenta una atenuación más rápida en la región de transición. Por otro lado y dado que se trata de una medida prácticamente en continua, los efectos de la distorsión de fase introducidos por el filtro paso bajo no afectan a la medida. Con estas consideraciones generales se ha diseñado un filtro activo paso bajo, de segundo orden con estructura Sallen-Key. Los valores de los componentes seleccionados, así como la respuesta frecuencial se muestran en la figura 4.39. Para el diseño se ha utilizado el programa FilterLab de Microchip Technology.

EJERCICIO 4.10

Considérese el sistema multiplexado de la figura 4.40 en el que se presentan cuatro canales analógicos de entrada, CH0 a CH3. Diseñe un posible filtro que responda adecuadamente cuando el multiplexor conmuta entre dos entradas analógicas de distinto nivel de continua.

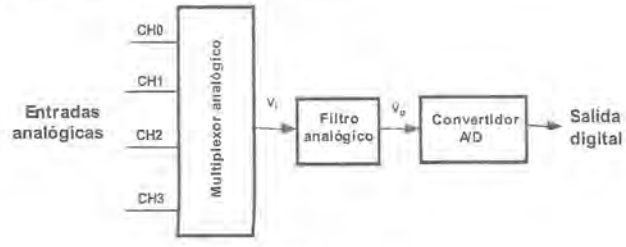


Figura 4.40

SOLUCIÓN:

El filtro ha de ser capaz de responder con suficiente rapidez cuando el multiplexor efectúe una conmutación de canal. Por tanto, el aspecto más importante a considerar será la velocidad de respuesta del filtro ante un escalón de tensión, ya que el convertidor A/D estará esperando por una entrada de tensión válida.

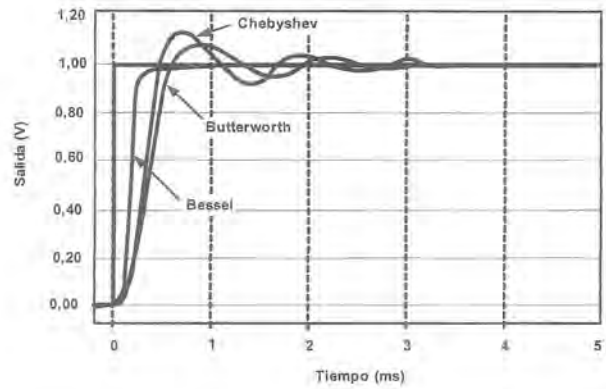


Figura 4.41

Para esta aplicación la aproximación de Chebyshev no es adecuada al presentar una excesiva sobreoscilación. La aproximación de Butterworth sería la mejor solución si no fuera por la sobreoscilación que presenta. La aproximación de Bessel tiene una buena respuesta ante un escalón de tensión, alcanzando el valor final de forma rápida. En cambio, su respuesta frecuencial en la banda de paso no es buena. Esto último se puede paliar aumentando el orden del filtro, a costa de un incremento de componentes. Estas consideraciones han llevado a diseñar un filtro paso bajo de cuarto orden, de Bessel con estructura Sallen-Key y frecuencia de corte de 1 kHz. Las figuras 4.41 y 4.42 muestran los resultados obtenidos siguiendo el método general de diseño expuesto en los apartados anteriores.

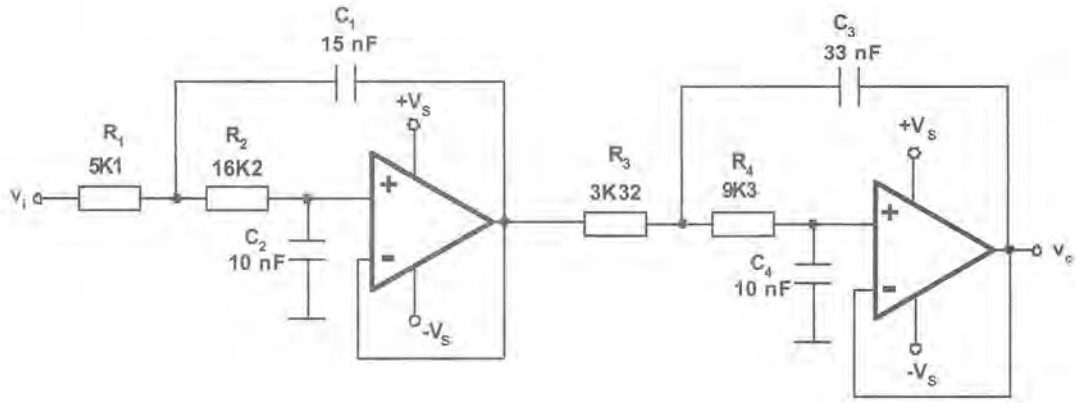


Figura 4.42

Resumen

- El filtrado es una parte esencial en el procesamiento de señales analógicas. Permite reducir el nivel de señales no deseadas, limitar el ancho de banda, facilitar la recuperación de señales o minimizar el *aliasing* en los sistemas de adquisición de datos.
- Los filtros pasivos emplean exclusivamente elementos R, L, C. Es frecuente encontrar en instrumentación filtros pasivos RC, si bien se trata de aplicaciones poco exigentes y donde la ganancia es la unidad. Los filtros pasivos con bobinas se emplean principalmente en aplicaciones de electrónica de potencia.
- Los filtros activos resuelven algunas de las limitaciones de los filtros pasivos, ofreciendo aislamiento entre etapas. Su empleo en alta frecuencia está limitado por las características dinámicas de los amplificadores operacionales.
- Existen diversas aproximaciones matemáticas a la hora de aproximar la característica del filtro. Cada una de ella trata de optimizar una determinada característica del filtro. La aproximación de Butterworth es la que presenta un mejor compromiso entre respuesta de ganancia y respuesta de fase. No tiene rizado en la banda de paso ni en la banda suprimida, a costa de una relativamente ancha región de transición.
- Además de la frecuencia de corte y el orden, en ocasiones es necesario especificar otras características del filtro como la frecuencia límite de la banda de paso, la frecuencia a la que comienza la banda suprimida, el rizado máximo de la banda de paso o la atenuación mínima de la banda suprimida.
- Una vez establecidas las especificaciones del filtro, el diseño del mismo se reduce a calcular los coeficientes de los polinomios característicos. Esto puede realizarse con la ayuda de tablas o bien mediante herramientas software, algunas de ellas proporcionadas de forma gratuita por los fabricantes de circuitos integrados.
- Los filtros activos de orden igual o superior a tres se construyen a partir de la conexión en cascada de secciones de primer y segundo orden.
- Las estructuras físicas más comunes para implementar filtros activos son Sallen-Key, realimentación múltiple y variables de estado. La estructura Sallen-Key no invierte la fase de la señal de entrada, es la menos dependiente de las características del operacional. Suele emplearse en filtros de ganancia unidad y para valores de Q pequeños.
- Los filtros de variables de estado son los que permiten obtener una mayor precisión a costa de requerir un mayor número de componentes. Es posible un ajuste independiente de todos los parámetros del filtro.
- Los filtros de capacidad conmutada se caracterizan por muestrear la señal analógica de entrada. Las resistencias son implementadas en silicio mediante interruptores CMOS y condensadores, por lo que el consumo es mínimo. Además, es posible ajustar la frecuencia de corte del filtro con una gran precisión, ya que su valor depende de la frecuencia de reloj y de una relación de capacidades.
- En la implementación de un filtro hay que tener en cuenta las características reales de los componentes tanto activos como pasivos, pues de ellos va a depender la calidad de la respuesta del filtro. En general los valores de las resistencias estarán en el margen de 100 Ω a 1 MΩ y serán preferiblemente de película metálica, en montaje superficial y con una tolerancia máxima del 1%.
- En cuanto a los condensadores, se deben evitar valores de capacidad inferiores a 100 pF, preferiblemente cerámicos multicapa de montaje superficial, con dieléctricos NPO o X7R y tolerancias a ser posible del 1%.

Problemas propuestos

4.1 La figura 4.43 muestra un filtro paso bajo al que se ha conectado un circuito amplificador no inversor. Dibuje el diagrama de Bode (módulo y fase) del conjunto.

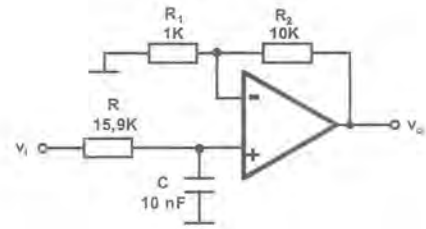


Figura 4.43

4.2 Para el filtro paso banda de la figura 4.44 determine: a) la función de transferencia; b) los valores de los componentes para obtener una frecuencia de corte inferior de 10 kHz, una frecuencia de corte superior de 100 kHz y una ganancia en la banda pasante de 60 dB.

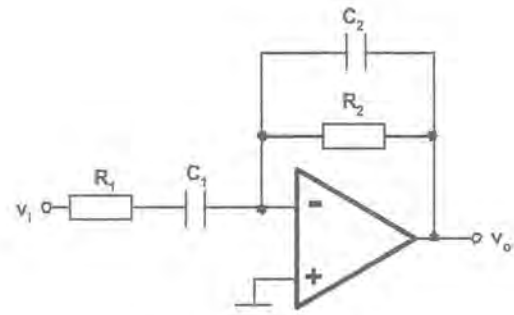


Figura 4.44

4.3 Determine el orden mínimo de un filtro RC paso alto de manera que su ganancia sea de 40 dB cuando la frecuencia sea $0{,}5f_c$, siendo f_c la frecuencia de corte del filtro.

4.4 Diseñe un filtro paso bajo que cumpla las siguientes especificaciones: ganancia en continua 0 dB, frecuencia de corte 10 kHz, Q = 0,5 y atenuación en alta frecuencia de 40 dB/década.

4.5 Conectando etapas Sallen-Key en cascada diseñe un filtro paso bajo de Butterworth que tenga las siguientes especificaciones: ganancia en continua 0 dB, frecuencia de corte 1 Hz y atenuación en alta frecuencia de 60 dB/déc.

4.6 Un sistema de medida óptico incluye un LED de infrarrojos y un fototransistor. El LED está excitado por una señal cuadrada de 1 kHz. Se desea diseñar un filtro analógico que elimine la señal de 100 Hz proveniente del alumbrado convencional, a fin de que no se sature el sistema de medida. Suponga que la ganancia deseada a 1 kHz es de 20 dB.

4.7 Para la medida de luz mediante un fotodiodo se emplea el circuito convertidor corriente-tensión de la figura 4.45. Razone qué tipo de filtro puede ser adecuado a fin de eliminar el ruido de alta frecuencia.

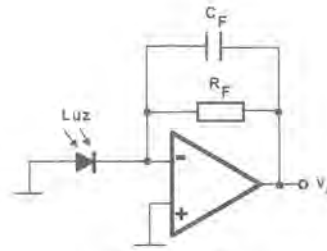


Figura 4.45

4.8 Para la medida de la presión arterial se utiliza un sensor de presión que proporciona una señal como la indicada en la figura 4.46a. A partir de esta señal se desea realizar un filtro que permita extraer, de la zona recuadrada, la señal de oscilación que se muestra en la figura 4.46b. Esta señal será posteriormente procesada para obtener el valor de la presión sistólica y diastólica.

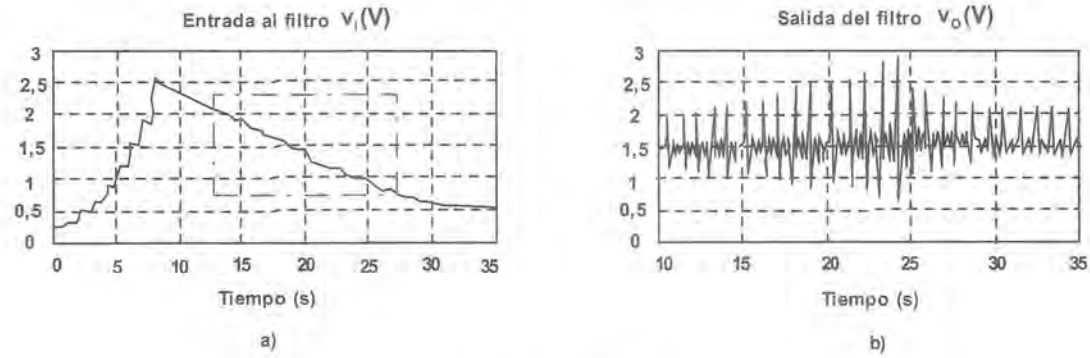


Figura 4.46

4.9 Un filtro paso banda tiene una frecuencia central de 1 kHz y un ancho de banda de 100 Hz. Determine: a) el valor del factor de calidad; b) las frecuencias de corte superior e inferior.

4.10 El filtro de la figura 4.47 se conoce como filtro paso todo. Su función de transferencia tiene una ganancia constante en todo el margen de frecuencias pero introduce un cierto desfase. Se pide: a) deducir la función de transferencia del filtro; b) dibujar el diagrama de Bode; c) indicar una posible aplicación del filtro.

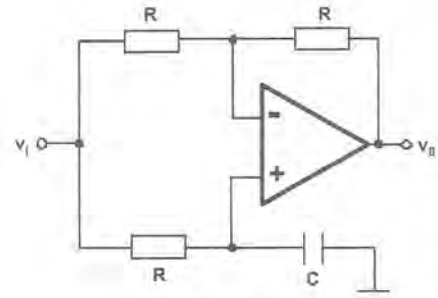


Figura 4.47

5

Sensores potenciométricos

5.1 Características generales
5.2 Parámetros característicos de los potenciómetros
5.3 Tipos de potenciómetros
5.4 Acondicionamiento de señal en potenciómetros
5.5 Errores debidos al cableado

5.1 Características generales

Un **potenciómetro** es un dispositivo electromecánico que consta de una resistencia de valor fijo sobre la que se desplaza un contacto deslizante, el **cursor**, que la divide eléctricamente. En la figura 5.1 se representa su símbolo eléctrico.

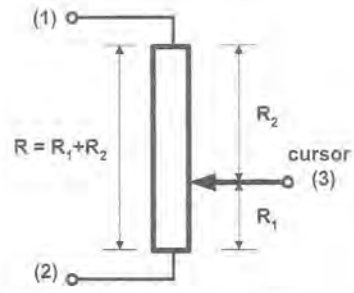


Figura 5.1. Símbolo eléctrico de un potenciómetro. R representa la resistencia de valor fijo, la punta de flecha señala el cursor o elemento deslizante. Entre el cursor y los extremos se miden las resistencias R_1 y R_2.

La figura 5.2 muestra la estructura de un potenciómetro en sus dos configuraciones más típicas. La resistencia fija adopta la forma de una pista de material resistivo. La pista puede ser recta, como muestra la figura 5.2a, o circular (figura 5.2b). El cursor es una pieza móvil que se desliza por medio de un eje sobre la pista resistiva describiendo movimientos lineales o angulares.

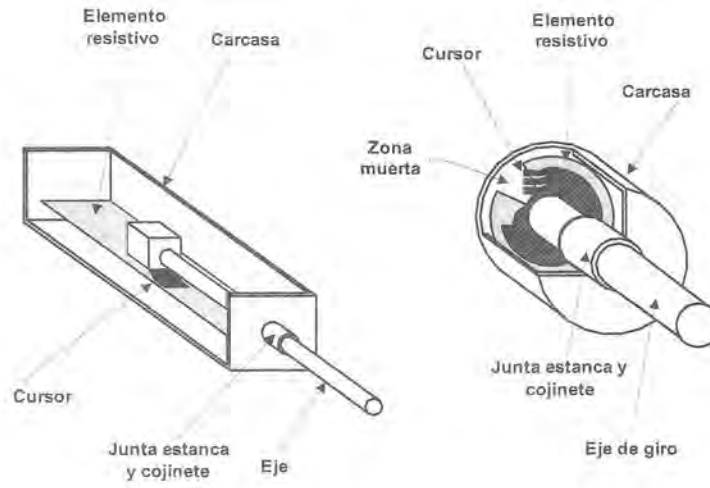


Figura 5.2. Estructura de un potenciómetro con (a) movimiento lineal, y (b) movimiento angular.

El cursor suele estar formado por dos o más escobillas de diferente longitud para prevenir los rebotes o las resonancias mecánicas originadas por movimientos rápidos o vibraciones. El conjunto se encapsula en una caja plástica o metálica que lo aísla del exterior.

La aplicación más común de este dispositivo en instrumentación es como sensor de desplazamiento de tipo resistivo. El movimiento del cursor origina un cambio en la resistencia medida entre el terminal central y uno cualquiera de los extremos (divisor resistivo). Este cambio de resistencia puede utilizarse para medir desplazamientos lineales o angulares de una pieza acoplada al cursor. Para la obtención de una señal eléctrica relacionada con el desplazamiento se alimenta, en general, con una tensión continua, adoptando la configuración de un simple divisor de tensión. A veces, esta es la única clase de acondicionamiento que se requiere (figura 5.3). Analizaremos en detalle esta configuración posteriormente.

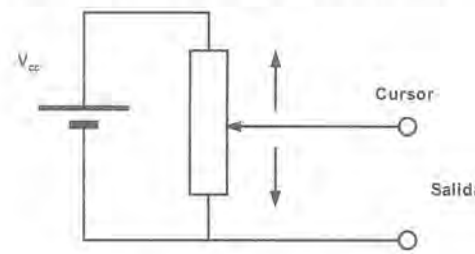


Figura 5.3. Potenciómetro para la medida de desplazamientos configurado como un divisor de tensión.

Los potenciómetros pueden utilizarse para medir diversas magnitudes físicas siempre que sea posible convertirlas en un desplazamiento. En la figura 5.4 se muestra, como ejemplo, un sensor de presión. El mismo principio puede utilizarse para medir otro tipo de magnitudes.

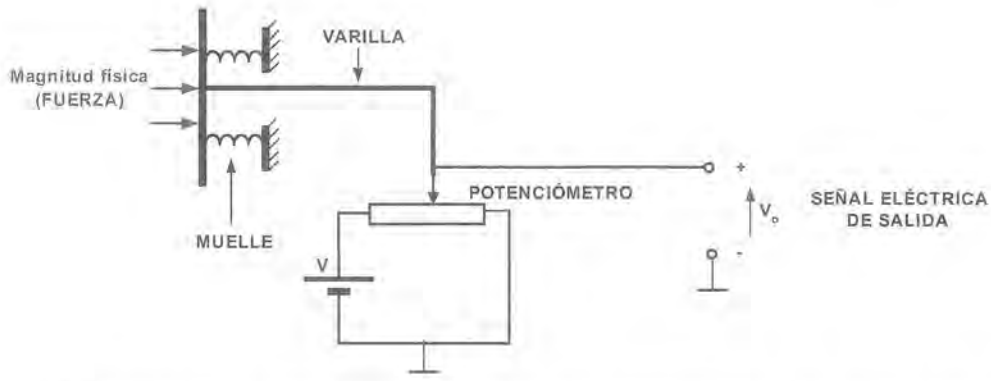


Figura 5.4. Ejemplo de un sensor de presión basado en un potenciómetro. Una fuerza aplicada sobre una superficie, presión, desplaza un muelle. El desplazamiento del muelle es proporcional a la fuerza ejercida y se lleva al cursor del potenciómetro mediante una varilla. La señal de tensión se obtiene del cursor del potenciómetro que está configurado como divisor de tensión.

A pesar de que estos sistemas fueron de los primeros que se emplearon en la instrumentación industrial, hoy en día, la aplicación de los potenciómetros como sensor de desplazamiento ha caído en desuso por factores relacionados con la complejidad mecánica que requieren sus aplicaciones —anclajes, rodamientos, guías— y por la necesidad de mantenimiento debido al desgaste que sufren sus partes móviles. En la actualidad, están siendo reemplazados por otros dispositivos más fiables basados en procedimientos de detección ópticos, como los codificadores de impulsos.

Además de como sensor, los potenciómetros tienen otras aplicaciones; por ejemplo, se utilizan también como resistencias de ajuste, de control, como reostato o como divisor de corriente. Los **potenciómetros de ajuste** (*trimmers*) se emplean para reducir el efecto de tolerancias en los circuitos y los potenciómetros de control se utilizan para variar el nivel de una señal.

5.2 Parámetros característicos de los potenciómetros

A continuación, se indican los parámetros fundamentales que caracterizan un potenciómetro.

Resistencia nominal: es la resistencia medida entre los terminales extremos cuando el cursor se encuentra desplazado al límite en un extremo[1]. La tolerancia del valor de esta resistencia suele ser alta. Los valores típicos son del ±10% al ±20%.

Resistencia de contacto concentrada (CCR): es el valor de la resistencia entre el contacto del cursor y la pista resistiva. Se calcula como:

$$CCR = \frac{r_{1-2} + r_{2-3} - r_{1-3}}{2r_{1-3}} \; 100\ [\%]$$

donde r_{i-j} es la resistencia entre los terminales i y j (1 y 3 son los terminales extremos).

[1] En algunos potenciómetros, como los bobinados, el cursor puede cortocircuitar varias espiras, modificando la resistencia medida entre los terminales extremos.

- **Variación de la resistencia de contacto (CRV):** es el cambio máximo instantáneo de la resistencia entre el contacto del cursor y la pista resistiva cuando se mueve el cursor. Este parámetro es importante para determinar el comportamiento dinámico de la resistencia de contacto.
- **Resistencia terminal:** es la resistencia medida entre el cursor y un extremo, con el cursor situado sobre dicho extremo.
- **Linealidad:** generalmente se especifica la linealidad independiente (véanse características estáticas en el capítulo 1). Se expresa como porcentaje del fondo de escala (%FS).
- **Potencia nominal:** potencia máxima que puede disipar. Se especifica a una temperatura determinada y permite calcular el valor máximo de tensión que puede aplicarse al potenciómetro.
- **Resolución:** en los potenciómetros de hilo bobinado la resolución está cuantificada al moverse el cursor entre las espiras del hilo. Se especifica mediante el inverso del número de espiras, N, expresado en tanto por ciento: $R = (1/N)\cdot 100$. Otros potenciómetros están formados por una pista continua de material resistivo y entonces la resolución se considera "ilimitada", aunque en la práctica está limitada por factores como: el ruido de la resistencia de contacto, la estabilidad de la tensión de alimentación, etc.
- **Ruido:** se cuantifica en términos de la resistencia parásita que aparece entre el cursor y la superficie resistiva al desplazarse el cursor.
- **Coeficiente de temperatura:** establece el cambio en el valor de la resistencia del potenciómetro en función de la temperatura, se expresa como $\alpha = 1/R\ (dR/dT)$ donde T es la temperatura y R es la resistencia nominal del potenciómetro. Lo ideal es mantener este parámetro lo más bajo posible utilizando para ello materiales de bajo coeficiente térmico. Habitualmente se expresa en partes por millón (ppm).
- **Coeficiente de disipación:** cuantifica la capacidad del potenciómetro para la evacuación del calor generado por efecto Joule. Tiene unidades mW/°C y resulta un parámetro básico para conocer el incremento de temperatura que se produce en el potenciómetro cuando lo recorre una determinada corriente.
- **Resistencia térmica:** R_θ, es el valor inverso del coeficiente de disipación.
- **Resistencia de aislamiento:** resistencia entre cada terminal y las superficies conductoras externas del potenciómetro para una tensión determinada.
- **Rotación/recorrido eléctrica/o y mecánica/o:** en los potenciómetros no todo el recorrido del cursor está recubierto por material resistivo. La porción recubierta por material resistivo define la rotación o recorrido eléctrico efectivo; la total define la mecánica.
- **Vida rotacional útil:** mide el número de ciclos que soporta antes de deteriorarse.
- **Par resistente:** expresa la oposición al avance del cursor por el rozamiento con la superficie resistiva.

5.3 Tipos de potenciómetros

Los potenciómetros pueden clasificarse atendiendo a diversos criterios: según el elemento resistivo utilizado, por el tipo de desplazamiento que realice el cursor o por su función característica, es decir, por cómo cambie la resistencia en función del desplazamiento.

El elemento resistivo es una parte esencial en cualquier potenciómetro hasta el punto que afecta a todas sus características eléctricas. La clasificación más general distingue entre:

- Potenciómetros de hilo bobinado (*wirewound*).
- Potenciómetros no bobinados (*nonwirewound*):
  - Cermet.
  - Carbón.
  - Plástico conductivo.
  - Película metálica.
  - Híbridos.

En los potenciómetros de hilo bobinado el elemento resistivo es un hilo arrollado sobre un soporte. En los no bobinados el elemento resistivo es una pista de diversa composición (figura 5.5). Los no bobinados pueden a su vez clasificarse en: cermet (aleación de cerámica y metal **cer**ámica+**met**al), carbón, plástico conductivo (película de carbón), de metal y de película metálica. También es posible realizar combinaciones para mejorar ciertos parámetros eléctricos, dando lugar a los potenciómetros denominados híbridos. Por ejemplo, puede combinarse hilo bobinado con plástico conductivo y también cermet con plástico conductivo.

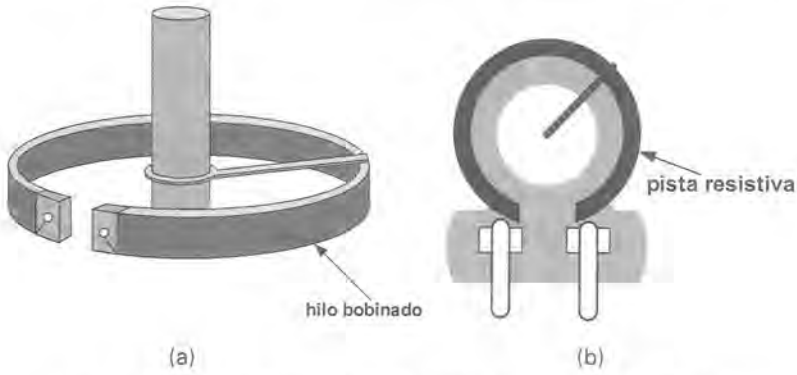


Figura 5.5. Potenciómetros: (a) bobinados, (b) no bobinados.

En general, en la fabricación de los potenciómetros se buscan unas características ideales. Así, el sustrato debe poseer la suficiente calidad para que no se vea afectado por las condiciones ambientales: deberá resistir la oxidación y evitar reacciones químicas que originen contactos pobres entre el cursor y la pista conductiva, lo que constituye una fuente de ruido. Tampoco deberá cambiar la resistencia con la tensión aplicada —esto ocurre con los potenciómetros de carbón—, y sus características no deben variar con el tiempo, es decir, debe tener bajas derivas (*drift*). También es importante que el coeficiente térmico de la resistencia sea pequeño para que presente baja sensibilidad ante cambios de la temperatura. La resistencia y la fricción entre el cursor y el sustrato también debe ser pequeña con el objeto de evitar el desgaste y prolongar su vida útil. En relación con estos objetivos cada tipo presenta sus ventajas e inconvenientes:

- La característica esencial de los bobinados es su excelente estabilidad térmica junto con su capacidad para manejar niveles de potencia considerables. Generalmente, se utiliza un hilo compuesto por una de estas aleaciones: níquel-cromo, cobre-níquel y oro-platino. Su resolución, sin embargo, es limitada debido a las discontinuidades entre las espiras (figura 5.5a). Los valores resistivos de los potenciómetros bobinados suelen ser relativamente pequeños y tienen limitaciones cuando se utiliza con señales alternas debido a los parásitos inductivos y capacitativos asociados.

- Los potenciómetros de plástico conductivo son muy utilizados. El plástico conductivo se fabrica con resinas plásticas (epoxy, poliésteres o poliamidas) mezcladas con polvos de carbono que se depositan sobre un sustrato plástico o cerámico. Presentan una resolución "ilimitada", baja fricción y niveles de ruido también bajos, gracias al acabado en espejo de su película plástica. Esta película puede, incluso,

ajustarse mediante técnicas láser para conseguir el valor resistivo y la respuesta deseada. Su respuesta en frecuencia es también buena. Por contra, su capacidad de disipación no es muy alta, la corriente por el cursor deben limitarse a valores bajos y les afecta la temperatura moderadamente.

- En los híbridos se combinan las prestaciones de los de hilo bobinado y plástico conductivo. Este tipo de potenciómetro es muy utilizado en medidas que requieran cierta precisión, aunque el componente plástico limita la potencia que puede disipar.
- Los potenciómetros de cermet se fabrican combinando partículas muy finas de cerámica y metales preciosos que se sinterizan en un horno, resultando una película que se deposita sobre un sustrato cerámico. Los materiales utilizados y su proceso elaboración son muy diversos y cada fabricante suele utilizar sus propios compuestos que en ocasiones no divulga por razones competitivas. No obstante, el término cermet se aplica de forma genérica a todos ellos aunque no sean iguales en su composición. El material resistivo se comporta muy bien frente a la temperatura y la humedad. Su coeficiente de temperatura es muy bajo y disipa más potencia que los de plástico conductivo aunque su nivel de ruido y vida útil son peores. Su resolución es muy buena y también su estabilidad y respuesta en frecuencia. Su vida útil es excelente debido a que la superficie es muy dura, los fallos suelen venir por desgaste del cursor.
- Los potenciómetros de película metálica fina se fabrican depositando una película metálica muy fina por deposición de vapores metálicos o *sputtering* sobre un sustrato cerámico. Pueden utilizarse diversos metales aunque suele utilizarse una aleación de níquel-cromo. Su respuesta en frecuencia es muy buena y su ruido muy bajo. Los valores resistivos disponibles suelen ser bajos entre 10 Ω y 20 kΩ. Están algo limitados en potencia y su vida útil es algo corta.

  Los potenciómetros metálicos de película gruesa presentan como característica más relevante su coeficiente de temperatura que es extremadamente bajo (menos de 10 ppm/°C).
- En los potenciómetros de carbón se compactan polvos de grafito con resinas. Les afectan bastante los cambios de temperatura y la humedad debido a las resinas y la resistencia puede llegar a variar según la tensión aplicada. Su principal ventaja reside en su bajo coste comparado con el resto. Así, suelen utilizarse en aplicaciones de propósito general pero no para realizar medidas con precisión.

En el campo de la instrumentación suele utilizarse un tipo especial de potenciómetro denominado **potenciómetro de precisión**; suele fabricarse en hilo bobinado, plástico conductivo o híbrido. En general posee una alta calidad, pero su tamaño y coste es superior a los normales. Entre sus prestaciones destaca su capacidad para reducir la histéresis y el par resistente que ofrece al avance, además de soportar esfuerzos en más de una dirección.

La tabla 5.1 compara las características de los potenciómetros más utilizados en la instrumentación, y la tabla 5.2 muestra, a modo de ejemplo, los parámetros característicos de dos potenciómetros de distinta tecnología (plástico conductivo e hilo bobinado) donde pueden contrastarse sus prestaciones. Observe, por ejemplo, como el coeficiente de temperatura es muy bajo en un potenciómetro bobinado, jugando en su contra la resolución.

	Plástico conductivo	Hilo bobinado	Híbridos
Resolución	Infinita	Limitada	Infinita
Potencia	Baja	Alta	Baja
Estabilidad térmica	Pobre	Excelente	Muy buena
Ruido	Muy bajo	Bajo	Bajo
Vida media	10^6-10^8 ciclos	10^5-10^6 ciclos	10^6-10^7 ciclos
Comportamiento en alta frecuencia	Buena	Pobre	Buena

Tabla 5.1. Resumen de las características de los principales tipos de potenciómetros [5.1].

Parámetro	Potenciómetro de precisión (plástico conductivo)	Potenciómetro de precisión (hilo bobinado)
Pista	Plástico conductivo	Hilo bobinado
Tolerancia de la resistencia	±10%	±5%
Linealidad	± 1% (independiente)	± 1% (independiente)
Potencia nominal	1 W (de 0°C a 70°C)	2 W a 40°C
Rigidez dieléctrica	750 Vrms	1000 Vrms
Resistencia de aislamiento	1.000 MΩ	1.000 MΩ a 500 vdc
Rotación eléctrica	340° ±4°	320° ±5°
Rotación mecánica	340° ±5°	330° ±5°
Par de funcionamiento	1,8 g·cm	14,4 g·cm
Temperatura de funcionamiento	de –65°C a +125°C	de –55°C a +125°C
Resolución	infinita	limitada
Coeficiente de temperatura	±600 ppm/°C (max)	20 ppm/°C
Vida útil rotacional	5.000.000 ciclos	500.000

Tabla 5.2. Tabla resumen de los valores típicos de dos potenciómetros de precisión con sustrato de hilo bobinado y plástico conductivo.

Atendiendo al tipo de desplazamiento realizado por el cursor se clasifican en potenciómetros de:

- Movimiento lineal.
- Rotatorios:
  - Sencillos.
  - Multivuelta.
  - Infinitas vueltas.
- Cuerda, "yo-yo" o de cable.

En los potenciómetros de movimiento lineal el cursor describe desplazamientos en línea recta como el tipo mostrado en la figura 5.2a. El desplazamiento puede ir desde milímetros hasta varios metros. Los rotatorios pueden ser sencillos o de una sola vuelta (figura 5.2b), o multi-vuelta, generalmente de 3, 5, 10 ó 15 vueltas. En los potenciómetros de una vuelta el cursor se mueve sólo en una sección de la circunferencia ya que tiene topes al final del recorrido. En los multi-vuelta el desplazamiento angular se convierte mediante un tornillo sin-fin en un desplazamiento lineal (figura 5.6a). Los de infinitas vueltas son similares a los de una vuelta pero se han suprimido los topes. Los potenciómetros suelen tener denominaciones estandarizadas para sus terminales. Una denominación muy habitual es la de C, CW y CCW que significan: cursor, terminal al que se aproxima el cursor cuando gira en el sentido horario y terminal del que se aleja el cursor cuando gira en el sentido horario (figura 5.6a).

Los de cuerda también denominados "yo-yo", o de cable, permiten medir la posición y la velocidad usando un cable flexible arrollado en una bobina que está sometida a la tracción de un muelle (figura 5.6b). Este tipo de potenciómetros llegan a superar varias decenas de metros midiendo la longitud de la cuerda extendida.

Finalmente, cabe considerar como criterio de clasificación su función característica, f(l), es decir, cómo cambia la resistencia en función del desplazamiento, l. Aunque puede utilizarse virtualmente cualquier tipo de función para f(l) las más habituales son funciones lineales, logarítmicas y antilogarítmicas que dan lugar a:

- Potenciómetros lineales.
- Potenciómetros logarítmicos o antilogarítmicos.
- Otras funciones: seno, tangente, etc.

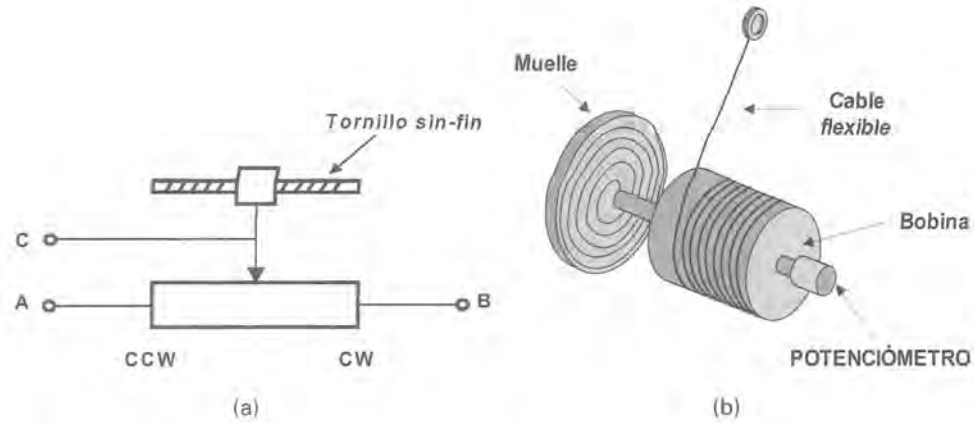


Figura 5.6. (a) Potenciómetro multivuelta, (b) potenciómetro de cable o yo-yo.

Los potenciómetros lineales son aquellos en los que la función característica es lineal, es decir, la resistencia del potenciómetro es constante por unidad de longitud.

El uso de los potenciómetros logarítmicos y antilogarítmicos no suele ser habitual en el campo de la instrumentación electrónica. Se utilizan, por ejemplo, en los equipos de sonido como control de volumen para tratar de acomodar el desplazamiento realizado con el nivel de sonido percibido, pues el oído humano tiene una respuesta logarítmica.

5.4 Acondicionamiento de señal en potenciómetros

La configuración de un potenciómetro como un simple divisor de tensión proporciona una forma de acondicionamiento sencilla para medir desplazamientos lineales o angulares, como se sugirió con anterioridad (figura 5.3). Si suponemos que la resistencia del potenciómetro es lineal, es decir, su resistencia es constante por unidad de longitud, la tensión de salida es proporcional al desplazamiento realizado. A continuación, justificaremos este resultado y veremos después el efecto que tiene en el circuito la conexión de una carga.

Consideremos el circuito de la figura 5.7a formado por un potenciómetro lineal de resistencia nominal R y de longitud de desplazamiento máxima L. Sea l el desplazamiento actual medido desde un punto tomado como referencia, 0, y sea x = l/L la fracción de desplazamiento recorrido. Como el potenciómetro es lineal, a una fracción de recorrido, x, le corresponde una fracción equivalente de la resistencia nominal R. Se obtiene así las relaciones:

$$R_1 = Rx$$

$$R_2 = R(1-x)$$

donde R_1 equivale a la porción de resistencia del potenciómetro comprendida entre 0 y el cursor y R_2 el resto, según se indica en la figura 5.7a.

El comportamiento del potenciómetro como elemento de un circuito depende de su impedancia de entrada y de su impedancia de salida. La impedancia de entrada, R_i y de salida, R_o, del potenciómetro sin carga, son:

$$R_i = R$$

$$R_o = R_1 \| R_2 = Rx\,(1-x)$$

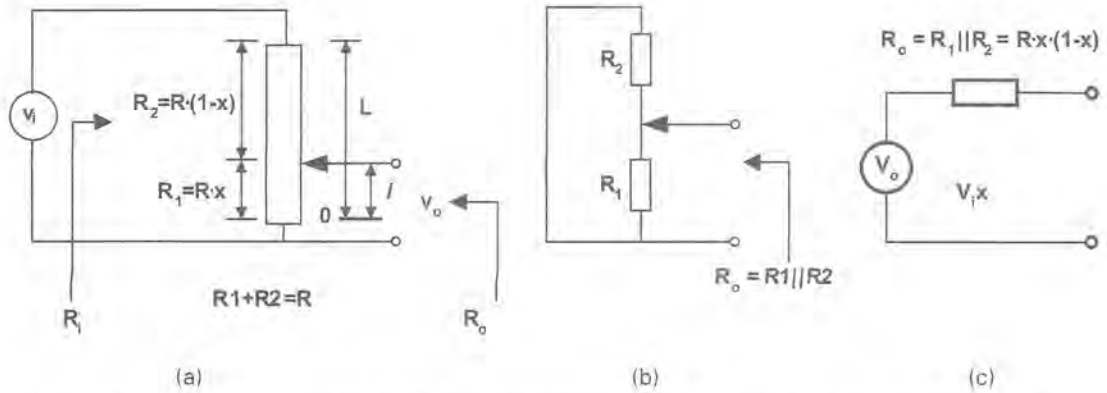


Figura 5.7. (a) Acondicionamiento de un potenciómetro para la medida de desplazamiento; (b) circuito utilizado para el cálculo de la impedancia de salida; (c) circuito equivalente de Thévenin visto entre el cursor y el extremo de referencia.

En la figura 5.7b se muestra el circuito para el cálculo de la resistencia de Thévenin o resistencia de salida donde la fuente de entrada se sustituye por su resistencia interna (se considera nula). En la figura 5.7c se muestra el circuito equivalente de Thévenin resultante visto desde los terminales de salida. En vacío, es decir, sin carga, la impedancia de entrada es constante, mientras que la impedancia de salida cambia en función de la fracción de desplazamiento x (figura 5.8).

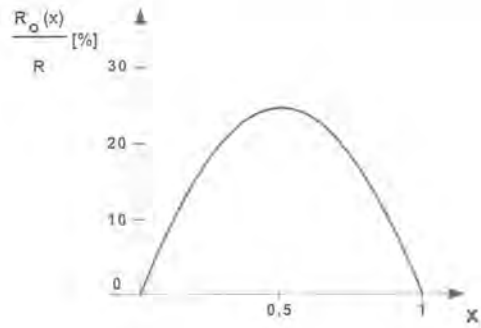


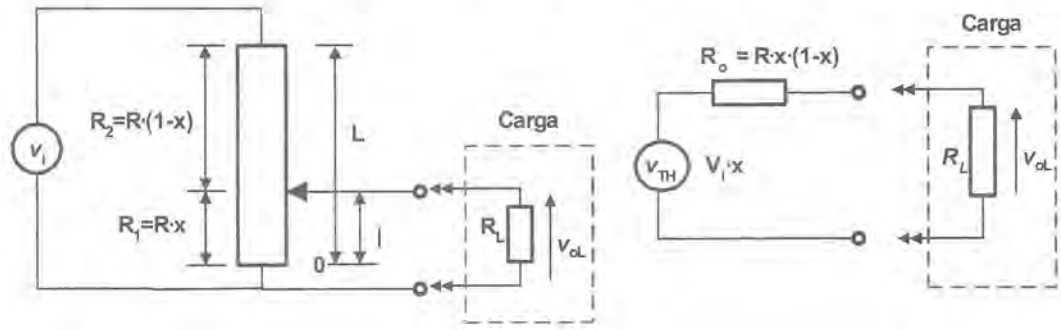
Figura 5.8. Impedancia de salida del potenciómetro en función del desplazamiento. El cambio en la impedancia de salida se expresa como porcentaje respecto a la resistencia nominal del potenciómetro.

Los cambios en la impedancia de salida no afectan a la respuesta del potenciómetro ya que no circula corriente por R_o al considerar la salida en circuito abierto. La tensión de salida, v_o, coincide con la tensión de Thévenin:

$$v_o = v_i \frac{R_1}{R_1 + R_2} = v_i x = v_i \frac{l}{L}, \qquad x \in [0,1] \tag{5.1}$$

Por tanto, la tensión de salida varía de forma lineal con el desplazamiento, l. Sin embargo, para que el circuito propuesto realice alguna función útil, es necesario conectar su salida a la etapa de entrada de otro dispositivo, en cuyo caso, la impedancia de entrada de este dispositivo se comportará como una carga para el potenciómetro.

En la figura 5.9a se representa la nueva situación considerando una carga. La tensión de salida se deriva ahora de un divisor de tensión entre R_o y R_L, tal como se muestra en el circuito equivalente de la figura 5.9b.



(a) (b)

Figura 5.9. (a) Sensor potenciométrico con carga, y (b) circuito equivalente de Thévenin.

En este circuito, el valor de R_o sí es importante ya que va a afectar a la tensión de salida y, por tanto, a la medida. Si pretendemos que toda la tensión de Thévenin se acople en la carga, la impedancia de salida debe anularse pero esto sólo ocurre en los extremos de desplazamiento del cursor (figura 5.8). Recuerde que R_o no es constante y llega a ser hasta un 25% del valor nominal del potenciómetro, justo en la mitad del recorrido. En el resto de posiciones, excepto en los extremos, el efecto de la impedancia de salida va a influir sobre la tensión leída, es decir, va introducir un error respecto a la respuesta ideal.

Una opción para paliar, en parte, el problema es hacer que la carga sea grande comparada con la impedancia de salida del potenciómetro. La impedancia de entrada también se modifica con el potenciómetro cargado, siguiendo una variación comprendida entre la resistencia nominal del potenciómetro (cursor en x = 0) y su paralelo con la carga (cursor en x = L). No obstante, en las condiciones habituales de trabajo esta variación sólo afecta a la propia fuente de alimentación. Si la fuente presenta una resistencia interna despreciable, el problema no es importante.

Analizando el circuito del potenciómetro con carga de la figura 5.9 se obtiene:

$$v_{oL} = v_i x \frac{R_L}{Rx(1-x)+R_L},$$

y dividiendo por R_L resulta:

$$v_{oL} = v_i x \frac{1}{\frac{x(1-x)}{k}+1}, \quad \text{con } k = R_L/R \quad \text{y} \quad x \in [0,1] \tag{5.2}$$

La relación entre la tensión de salida y el desplazamiento se muestra en la figura 5.10. Al conectar una carga al potenciómetro se introduce en el sistema un **error de no linealidad**. La magnitud de esta no linealidad depende de *k*, que es la relación entre la resistencia de carga y la nominal del potenciómetro.

Si R_L es considerablemente mayor que R, es decir, para un k suficientemente grande, la salida es prácticamente lineal como se aprecia en la figura 5.10 con k = 10. En general, puede tomarse para k un valor al menos de 10 (observe que el término x(1-x) en (5.2) está acotado en 0,25 y si k > 10 el resultado se desprecia frente al 1 del denominador). Si el valor de R_L es similar al valor de R, (k = 1) la salida presenta un error

de linealidad apreciable. En resumen, para conseguir una respuesta lineal se debe exigir al sistema que se acople una impedancia de entrada alta comparada con la resistencia nominal del potenciómetro.

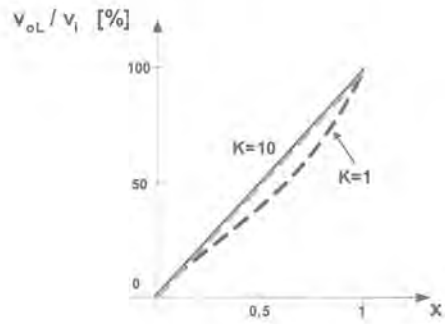


Figura 5.10. Salida del potenciómetro con carga, se expresa como % respecto a FS. Se aprecia el error de linealidad. La curva sólida representa la respuesta ideal, sin carga. Las curvas a trazos representan la respuesta para k = 1 y k = 10.

EJERCICIO 5.1

Obtenga una cota para el error absoluto del circuito de la figura 5.9a.

SOLUCIÓN:

El error absoluto se calcula como: e_a = |salida ideal (sin carga) — salida real (con carga)|

$$e_a = |v_o - v_{oL}| = v_i \frac{\frac{x^2(1-x)}{k}}{1+\frac{x(1-x)}{k}}, \text{ con } k = R_L/R$$

Para valores de k utilizados en la práctica, generalmente k>10, la expresión anterior se reduce a:

$$e_a \approx v_i \frac{x^2(1-x)}{k}$$

El error absoluto máximo se calcula resolviendo $de_a(x)/dx = 0$ y verificando que $d^2e_a(x)/dx^2 < 0$. Se obtiene un máximo en x = 2/3 por tanto:

$$e_{a(máx)} = e_a\left(\frac{2}{3}\right) = \frac{4}{27}\frac{v_i}{k},$$

que es la cota superior del error absoluto. El error absoluto referido al fondo de escala y expresado en porcentaje es:

$$e_{aFS} = \left|\frac{v_o - v_{oL}}{v_i}\right| \cdot 100$$

$$e_{aFS\,(máx)} = \frac{400}{27}\frac{1}{k}\,\% = \frac{15}{k}\%, \tag{5.3}$$

que es la cota superior del error absoluto referido a FS [%].

EJERCICIO 5.2

Obtenga una cota para el error relativo del circuito de la figura 5.9a.

SOLUCIÓN:

El error relativo es:

$$e_r(x) = \left| \frac{v_o - v_{oL}}{v_o} \right|,$$

sustituyendo resulta:

$$e_r(x) = \frac{v_i}{v_o}\left[\frac{x^2(1-x)/k}{1+x(1-x)/k}\right] = \frac{1}{x}\left[\frac{x^2(1-x)/k}{1+x(1-x)/k}\right] = \frac{x(1-x)/k}{1+x(1-x)/k}$$

Para valores suficientemente grandes de k, esta expresión se reduce a:

$$e_r \approx \frac{x(1-x)}{k} = \frac{x-x^2}{k}$$

El error relativo máximo se calcula de forma idéntica al caso anterior obteniendo un máximo para $x = 0{,}5$

$$e_{r(máx)} = e_r(0{,}5) = \frac{0{,}25}{k}$$

que expresado en porcentaje resulta:

$$e_{r(máx)\%} = \frac{25}{k}\ [\%] \qquad (5.4)$$

lo que proporciona la cota superior del error relativo expresado en porcentaje.

La sensibilidad del circuito de la figura 5.9a puede estimarse a partir de:

$$S = \frac{dv_{oL}}{dx} \approx v_i, \quad \text{para} \quad k > 10 \qquad (5.5)$$

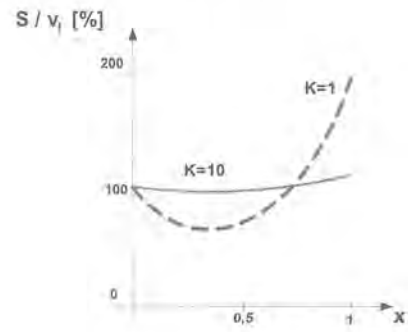


Figura 5.11. Variación de la sensibilidad del potenciómetro según relación de carga, k = 1 y k = 10. La sensibilidad se expresa como % respecto a la sensibilidad ideal sin carga (v_i). Observe que para k = 1 se produce un cambio considerable en la sensibilidad llegando a ser el doble en el extremo.

Por tanto, cuanta mayor tensión de alimentación mayor sensibilidad; sin embargo, este valor tiene un límite que impone potencia máxima admitida por el potenciómetro. La figura 5.11 muestra la variación de la sensibilidad para k = 1 y k = 10. El cambio acentuado de la sensibilidad para k = 1 expresa, en el fondo, la falta de linealidad.

EJERCICIO 5.3

Se desea medir el desplazamiento de una pieza en un margen de 6 cm mediante un potenciómetro. La salida del sistema se conecta a un medidor con una impedancia de entrada de 100 KΩ. Calcule el valor nominal del potenciómetro de forma que el error de linealidad máximo no exceda del 1%. El sistema diseñado deberá poseer la máxima sensibilidad teniendo en cuenta que los potenciómetros disponibles están limitados para soportar una potencia de 0,25 W como máximo.

SOLUCIÓN:

Como el error de no linealidad máximo debe ser inferior al 1% y de (5.3) resulta:

$15/k < 1$ con lo que $k = R_L/R > 15$, sustituyendo resulta $R < R_L/15 = 100\ k\Omega / 15 \approx 6{,}6\ k\Omega$ por lo que seleccionamos un potenciómetro de resistencia nominal $R = 6k\Omega$. Para obtener la máxima sensibilidad podemos aumentar la tensión de alimentación; sin embargo, la potencia máxima admitida por el potenciómetro limita este valor, así: $v_i^2/R < 0{,}25$ W con lo que $v_i < \sqrt{0{,}25 \cdot R} = 38{,}7$ V.

Si elegimos un valor de 30 V, se obtendrá una sensibilidad 30 V / 6 cm = 5 Vcm^{-1}.

Para reducir el error de linealidad pueden utilizarse varias estrategias. Una alternativa pasa por utilizar exclusivamente elementos pasivos como situar una resistencia en paralelo con una de las ramas del potenciómetro de valor igual a la carga [5.2]. Otras técnicas recurren a elementos activos como un seguidor de tensión para acoplar las impedancias o al uso de una fuente de alimentación simétrica.

5.5 Errores debidos al cableado

Además del error debido a la carga, otra fuente de error en los sistemas de medida basados en potenciómetros es el cableado: considere la situación descrita en la figura 5.12, donde el sensor se encuentra alejado del sistema que procesa la medida.

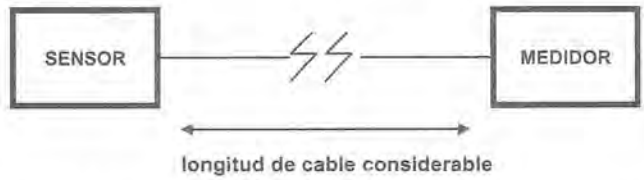


Figura 5.12. Potenciómetro alejado del sistema de medida.

En general, un cable presenta una resistencia muy pequeña, pero la situación cambia cuando supera cierta longitud. La situación es equivalente a la que se indica en la figura 5.13.

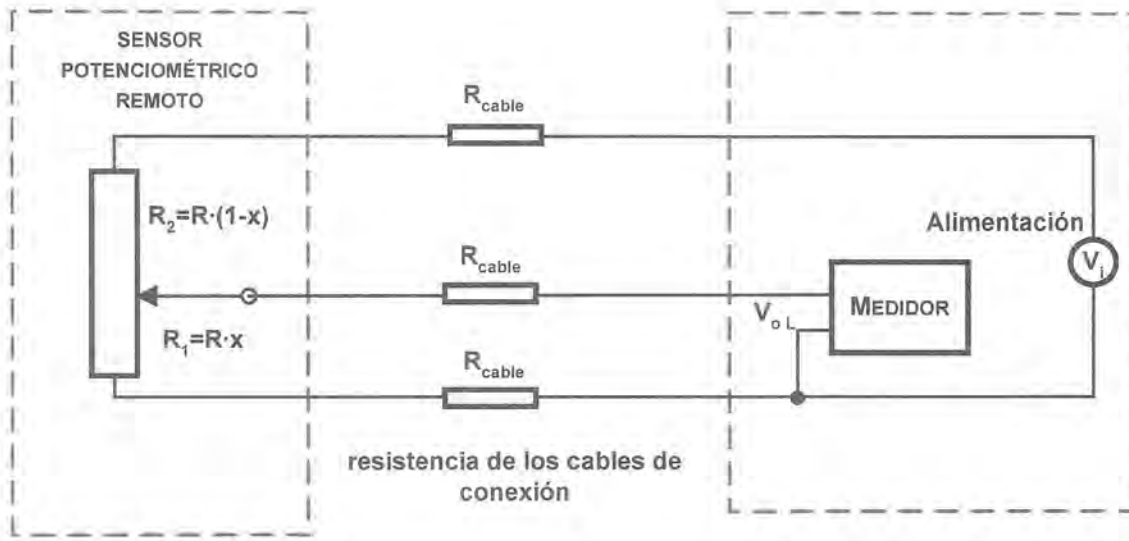


Figura 5.13. Sistema de medida a 3 hilos.

La resistencia de los cables de conexión se representa por R_{cable} y la señal de salida del potenciómetro se envía a través de estos cables al punto donde se procesa la medida. La alimentación también se envía de forma remota. Despreciando el error por carga, es decir, considerando que la impedancia de entrada del medidor es muy elevada comparada con la impedancia de salida del conjunto sensor potenciométrico-cables, se puede despreciar la corriente que circula a la entrada del medidor, resultando:

$$v_{oL} = v_i \frac{Rx + R_{cable}}{R + 2R_{cable}} = v_i \frac{x + h}{1 + 2h}, \tag{5.6}$$

que es la ecuación de una recta, siendo $h = R_{cable}/R$.

Del análisis de la ecuación anterior y de su comparación con el caso ideal se desprenden dos tipos de errores (figura 5.14). La primera observación es que $v_{oL} > 0$, es decir, se obtiene una tensión distinta de cero incluso para el caso en que $x = 0$. Este tipo de error se denomina **error de *offset*** o **error de cero**, su valor es:

$$v_{oL}(0) = v_i \frac{h}{1 + 2h} \tag{5.7}$$

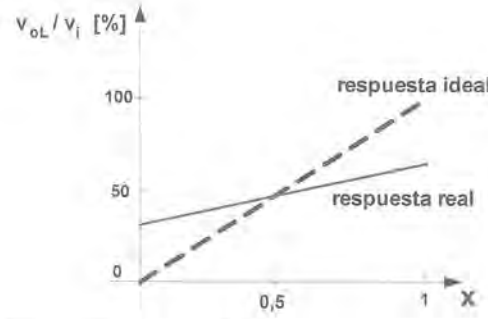


Figura 5.14. Error de *offset* y error de sensibilidad en una medida a 3 hilos. En línea a trazos se dibuja la respuesta ideal, en línea sólida la respuesta real con h = 1. La pendiente distinta indica el error en la sensibilidad.

El otro tipo de error tiene que ver con la sensibilidad. La sensibilidad se obtiene derivando (5.6) respecto a x:

$$S = v_i \frac{1}{1+2h} \tag{5.8}$$

La sensibilidad difiere de la ideal tal como se muestra por las pendientes de las curvas de la figura 5.14. Sólo para un h suficientemente pequeño se aproxima a la ideal.

Para eliminar el error de *offset* puede utilizarse un circuito de medida a 4 hilos (figura 5.15).

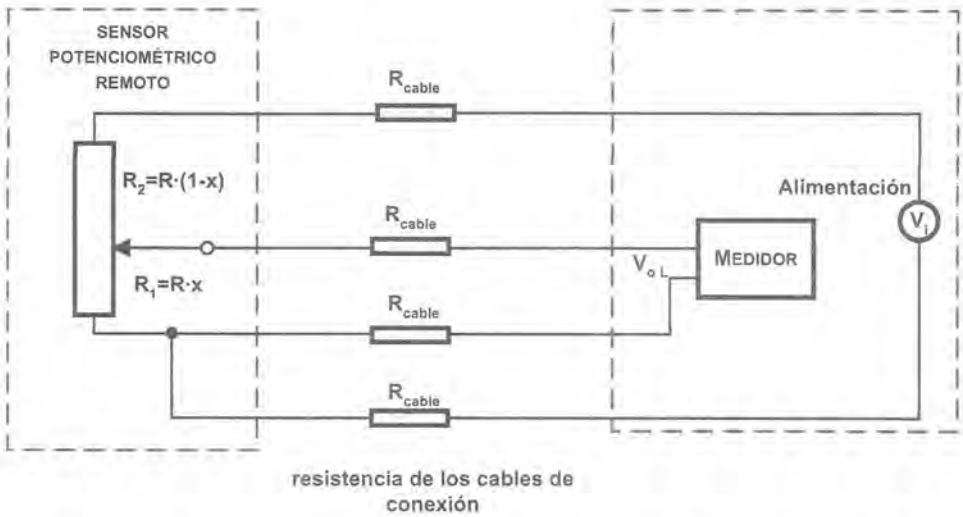


Figura 5.15. Circuito de medida a cuatro hilos para corregir el error de *offset*.

En este caso, suponiendo también despreciable el error por carga, se obtiene:

$$v_{oL} = v_i \frac{Rx}{R+2R_{cable}} = v_i \frac{x}{1+2h}, \qquad \text{siendo } h = R_{cable}/R \tag{5.9}$$

por lo que $v_{oL}(0) = 0$. La sensibilidad, sin embargo, no cambia, por lo que la técnica a 4 hilos corrige el error de cero aunque no se corrige el error de sensibilidad.

EJERCICIO 5.4

Un sensor potenciométrico bobinado de valor nominal 50 Ω se encuentra conectado remotamente a un medidor con una impedancia de entrada de 100 kΩ y a una fuente de alimentación de 3 V mediante un esquema de 3 hilos usando cobre de 0,5 mm² y 40 Ω/km. Calcule el error de offset que se comete si el sensor se encuentra a unos 150 m.

SOLUCIÓN:

La resistencia que representan los cables es:

$$150 \text{ m} \times 40\ \Omega/1000 \text{ m} = 6\ \Omega$$

$$h = (6\ \Omega / 50\ \Omega)$$

Aplicando la ecuación (5.7) resulta: $e_{offset} = 0{,}29$ V.

Resumen

Un **potenciómetro** es un dispositivo electromecánico que consta de una resistencia de valor fijo sobre la que se desplaza un contacto deslizante, el cursor, que la divide eléctricamente. La aplicación más común de este dispositivo en instrumentación es como sensor de desplazamiento de tipo resistivo aunque pueden utilizarse para medir otras magnitudes físicas siempre que sea posible convertirlas en un desplazamiento (fuerza, presión, etc.). En general, se une la pieza cuya posición se pretende detectar al cursor. Para la obtención de una señal eléctrica relacionada con el desplazamiento se alimenta por lo común con una tensión continua, adoptando la configuración de un divisor de tensión.

Existen diversos tipos de potenciómetros según el material que se utilice cada uno de ellos adaptados a una necesidad (potenciómetros de hilo bobinado, de película plástica, de cermet, etc.). En instrumentación suele utilizarse un tipo especial de potenciómetro denominado **potenciómetro de precisión** que presenta prestaciones especiales (baja histéresis, par resistente al avance reducido, capacidad para soportar esfuerzos en más de una dirección, etc.).

En el acondicionamiento es importante considerar los errores de linealidad que pueden introducirse por efecto de la carga y los errores debidos al cableado en situaciones de medida remota. Para minimizar los primeros, la carga debe poseer un valor mucho mayor que el nominal del potenciómetro. El error de *offset* introducido por el cableado puede corregirse utilizando técnicas de medida a cuatro hilos.

Además de como sensor, los potenciómetros tienen otras aplicaciones, por ejemplo, se utilizan también como resistencias de ajuste, de control, como reostato o como divisor de corriente.

Problemas propuestos

5.1 Se pretende conocer la posición de una pieza metálica que recorre una longitud máxima de 2 cm. Para ello, se utiliza un sensor potenciométrico cuya salida se aplica a la entrada de un convertidor A/D de 50 kΩ de impedancia de entrada. Los potenciómetros disponibles se encuentran limitados a 2 W. Proponer el sistema de medida teniendo en cuenta que debe dotarse al sistema de la máxima sensibilidad y el error de no linealidad no debe exceder del 1% del fondo de escala.

5.2 Calcular el error de *offset* y de sensibilidad de un sistema de medida a 3 hilos como el de la figura 5.13 si la longitud del cable es de 200 m y si se utiliza un cable de 20 Ω/km. Proponer un sistema para corregir ambos errores. Considere un potenciómetro de 100 Ω de resistencia nominal y una tensión de alimentación de 5 V.

5.3 ¿Por qué disminuye la sensibilidad en una medida remota a 3 ó 4 hilos utilizando un potenciómetro?

6

Sensores de temperatura de resistencia metálica

6.1 Características generales
6.2 Tipos de RTD. Comparación general
6.3 El autocalentamiento
6.4 Aplicaciones
6.5 Acondicionamiento de la señal

6.1 Características generales

Los metales se caracterizan por poseer coeficientes térmicos positivos de variación de la resistencia eléctrica; ello es producto de que al aumentar en los mismos la energía interna aumenta su resistividad.

Esta propiedad se utiliza desde hace mucho tiempo en el diseño y construcción de sensores de temperatura de resistencia metálica, que se designan con las siglas RTD (*Resistance Temperature Detector*). La figura 6.1 presenta el símbolo eléctrico de este dispositivo y el aspecto exterior de algunos sensores de temperatura de resistencia metálica comerciales.

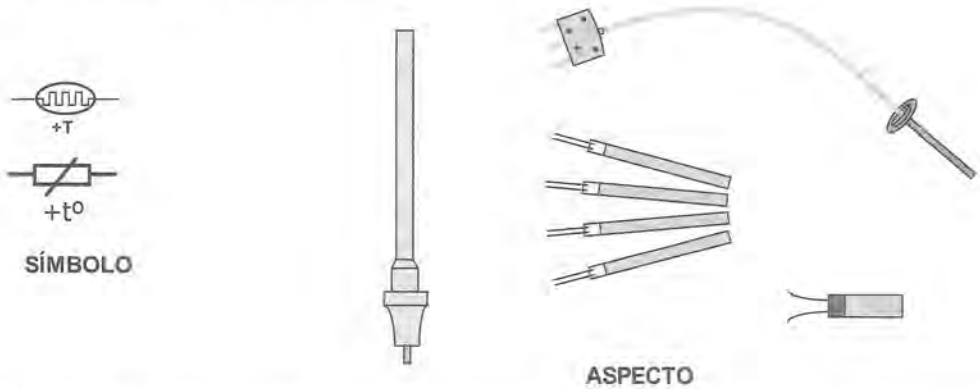


Figura 6.1. Símbolo y aspecto de sensores de temperatura de resistencia metálica (RTD) comerciales.

Si se tiene un conductor metálico de resistividad **ρ**, longitud **l** y área de la sección transversal **A**, su resistencia eléctrica vendrá dada por:

$$R = \frac{\rho \cdot l}{A} \tag{6.1}$$

Consideremos que un hilo metálico se encuentra inmerso en un medio a la temperatura T_a. Si el volumen y la masa del hilo son muy pequeños en comparación con los del medio, en el equilibrio térmico, el hilo adoptará la misma temperatura que el medio sin apenas "intercambiar" energía con el mismo, no afectando de esta manera al medio la inserción del hilo. En equilibrio térmico, el hilo presentará una resistencia R_{Ta} proporcional a la temperatura T_a. Si la temperatura del medio crece hasta $T_b > T_a$, la resistividad ρ e, incluso en alguna medida, la longitud l y el área A en el hilo cambian como producto de la dilatación del metal. Si consideramos los cambios en las dimensiones pequeños en comparación con el cambio de la resistividad ρ, entonces la resistencia del hilo R cambiará linealmente con la resistividad y ésta, a su vez, con la temperatura. Así, de (6.1) se desprende que ante un cambio en la temperatura del medio ΔT se producirá un cambio ΔR en la resistencia de la RTD linealmente dependiente de éste.

La tabla 6.1 presenta la resistividad y el coeficiente térmico de variación de la resistencia a la temperatura ambiente para algunos metales comúnmente empleados en la construcción de RTDs [6.1], [6.4].

Metales	Resistividad (ρ), [Ω·m]	Coeficiente térmico (α) [K^{-1}]
Platino, Pt	$10{,}6 \cdot 10^{-8}$	$3{,}9 \cdot 10^{-3}$
Níquel, Ni	$6{,}84 \cdot 10^{-8}$	$7 \cdot 10^{-3}$
Wolframio, W	$5{,}6 \cdot 10^{-8}$	$4{,}5 \cdot 10^{-3}$
Cobre, Cu	$1{,}68 \cdot 10^{-8}$	$4{,}3 \cdot 10^{-3}$

Tabla 6.1. Resistividad y coeficiente térmico de metales usados para construir RTDs.

Puede observarse que el platino es el metal, de los presentados en la tabla 6.1, que más bajo coeficiente térmico posee mientras que el níquel el de más alto. Si se construyen RTDs con estos metales, las de níquel serán las de mayor sensibilidad (mayor variación de resistencia ΔR ante un cambio en su temperatura ΔT), mientras que las de platino las de menor. Sin embargo, al poseer el platino la mayor resistividad puede conseguirse fabricar hilos muy finos con este metal con una apreciable resistencia sin necesidad de una gran longitud. En el níquel, los cambios de las dimensiones (que provocan cambios no lineales de la resistencia con la temperatura) son mucho más apreciables que en el resto de los metales por lo que ofrece una característica de resistencia en función de la temperatura con una gran no-linealidad. Este conjunto de propiedades y, fundamentalmente, la linealidad, es el que hace que no todos los metales sean utilizados para fabricar RTDs.

6.1.1 CURVAS DE CALIBRACIÓN

Los fabricantes ofrecen para cada RTD su tabla de calibración que no es más que una lista de los valores de resistencia R para cada temperatura T [6.2]. Con los datos de ésta, puede construirse la **curva de calibración** de la RTD, o sea, la gráfica de resistencia R_T de la RTD en función de su temperatura.

El valor de la resistencia R_T para la temperatura de cero grados centígrados se denomina R_0. La figura 6.2 presenta una parte de la curva de calibración de una RTD de platino conocida como Pt100 para la que el valor de la resistencia R_0 es 100 Ω.

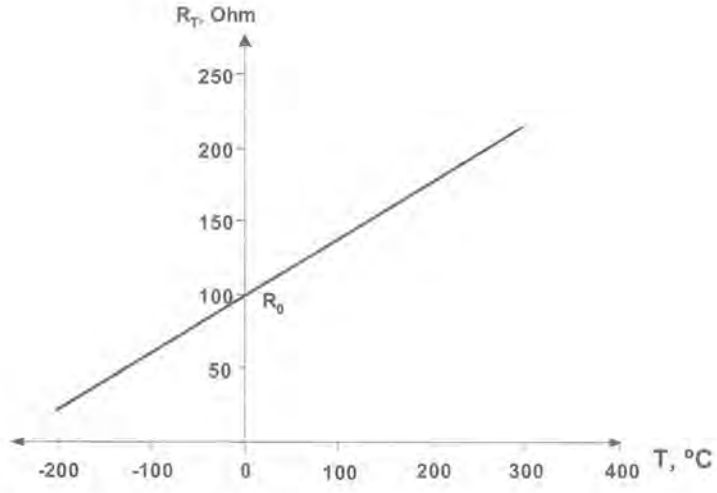


Figura 6.2. Curva de calibración de una Pt100 para $-200°C \leq T \leq +300°C$.

La figura 6.3 presenta las curvas de calibración normalizadas a R_0[1] para tres RTDs de diferentes materiales en el margen de –200 a +300°C. En este caso, el valor de R_0 es de 100 Ω en las de platino y cobre mientras que, en la de níquel, $R_0 = 120\ \Omega$.

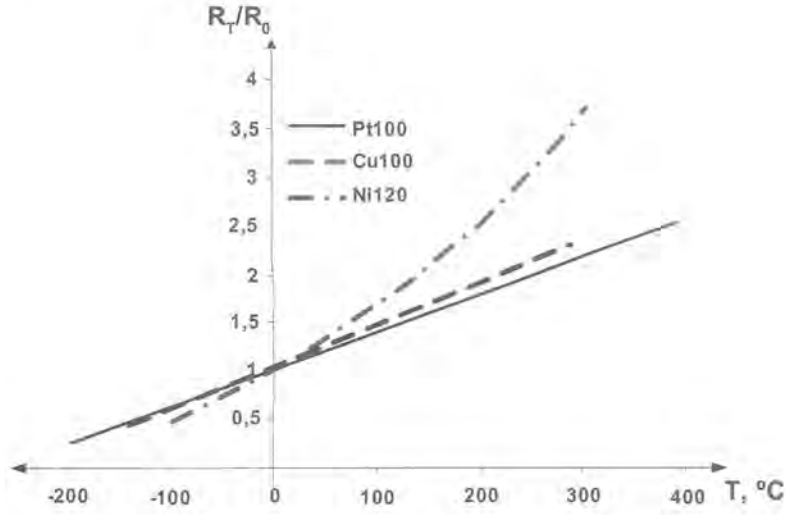


Figura 6.3. Curvas de calibración normalizadas de tres RTDs.

Se observa que la RTD de peor sensibilidad es la de platino, poco superada en este margen de temperaturas por la de cobre y muy superada por la de níquel. Pero si analizamos la linealidad de las curvas, la de níquel presenta una gran no-linealidad y por esta razón fundamental es por la que se prefiere utilizar las de platino o las de cobre en la mayoría de las aplicaciones.

[1] Las curvas de calibración normalizadas son muy útiles para comparar las respuestas de dispositivos ante una entrada común. La ordenada es el resultado del cociente de la resistencia R_T a cada temperatura T y R_0, por lo que aunque R_0 sea desigual en todas, se puede apreciar comparativamente la sensibilidad y linealidad.

6.1.2 MODELO MATEMÁTICO DE LA RTD

El modelo matemático que mejor se adecua a la curva de calibración de una RTD, desde el punto de vista estático, es el siguiente:

$$R_T = R_0 \cdot (1 + \alpha \cdot \Delta T + \beta \cdot \Delta T^2 + \gamma \cdot \Delta T^3 + ...) \quad (6.2)$$

donde: R_0 es la resistencia de la RTD a $T = T_0$ K,
ΔT es la variación de la temperatura $(T-T_0)$ y
α, β, γ... son los coeficientes del modelo para los términos lineal, cuadrático, cúbico, etc.

Por lo general, los valores de los coeficientes β y γ son muy pequeños[2] (del orden de 10^4 y 10^9 veces menores, respectivamente, que el valor de α para el platino) respecto al valor de α cuando se pretende aplicar (6.2) en un margen limitado de temperaturas por lo que, sin cometer un gran error para las aplicaciones más generales de instrumentación, el modelo de (6.2) se suele aproximar a:

$$R_T = R_0 \cdot (1 + \alpha \cdot \Delta T) \quad (6.3)$$

donde α es el coeficiente térmico de la RTD, K^{-1} y
R_0 es el valor de la resistencia de la RTD para T = 273 K = 0°C.

Como la magnitud del grado centígrado (°C) es igual al grado Kelvin (K), entonces (6.3) puede expresarse como

$$R_t = R_0 \cdot (1 + \alpha \cdot t) \quad (6.4)$$

donde R_0 es el valor de la resistencia de la RTD a t = 0°C, R_t es el valor de la resistencia de la RTD a la temperatura t, y t es la temperatura de la RTD en °C.

El comportamiento de la RTD ante variaciones en el tiempo de su temperatura se corresponde con el de un sistema lineal de primer orden de constante de tiempo τ, parámetro caracterizador de la respuesta dinámica de la RTD del cual depende el atraso en tiempo que existirá entre la variable temperatura de la RTD y la variación de su resistencia correspondiente al variar la frecuencia de la temperatura.

La constante de tiempo (τ) depende de muchos factores, siendo los fundamentales el tipo de material de la RTD, su masa, el área efectiva de termo-transferencia con el ambiente y la resistencia térmica existente entre el hilo metálico y el medio en el que se inserta.

Cuando la RTD es recubierta, para protección mecánica y aislamiento eléctrico respecto al medio, estos materiales aportan resistencias térmicas que hacen que el sistema pase a ser de segundo orden sobreamortiguado, pudiendo definirse el atraso en tiempo de la RTD en términos del tiempo de establecimiento o duración necesaria para que se alcance hasta el 99% o el 99,9% de la temperatura final para la RTD.

6.2 Tipos de RTD. Comparación general

Las RTDs pueden ser clasificadas atendiendo a los siguientes aspectos:

a) **Forma constructiva:** las RTDs pueden ser clasificadas en dos tipos, bobinadas y de rejilla o película metálica. La figura 6.4 presenta RTDs de platino fabricadas con estas tecnologías. En las bobinadas, el hilo conductor se bobina sobre un elemento aislante eléctrico y muy buen conductor del

[2] Para una RTD de platino en la que $\alpha = 3{,}9.10^{-3}K^{-1}$, $\beta = -5{,}775.10^{-7}K^{-2}$ y $\gamma = -4{,}183.10^{-12}K^{-3}$.

calor (cerámica). Este bobinado permite las dilataciones del hilo al cambiar la temperatura. La bobina es empaquetada con mica y colocada dentro de un tubo que le sirve de protección. En las RTDs de bobina suspendida el hilo metálico se bobina helicoidalmente dentro de tubos de cerámica que son rellenos con materiales cerámicos en polvo y sellados con vidrio en sus extremos. Por su parte, en las RTDs de rejilla metálica el hilo se dispone en un plano, permitiendo en este caso trabajar en la medida de temperatura de superficies. Una variante actual, de muy bajo coste y prestaciones similares es la de película metálica, en la que una fina capa de platino se deposita sobre un sustrato cerámico, cubierta por una película de vidrio fundido que sirve de protección.

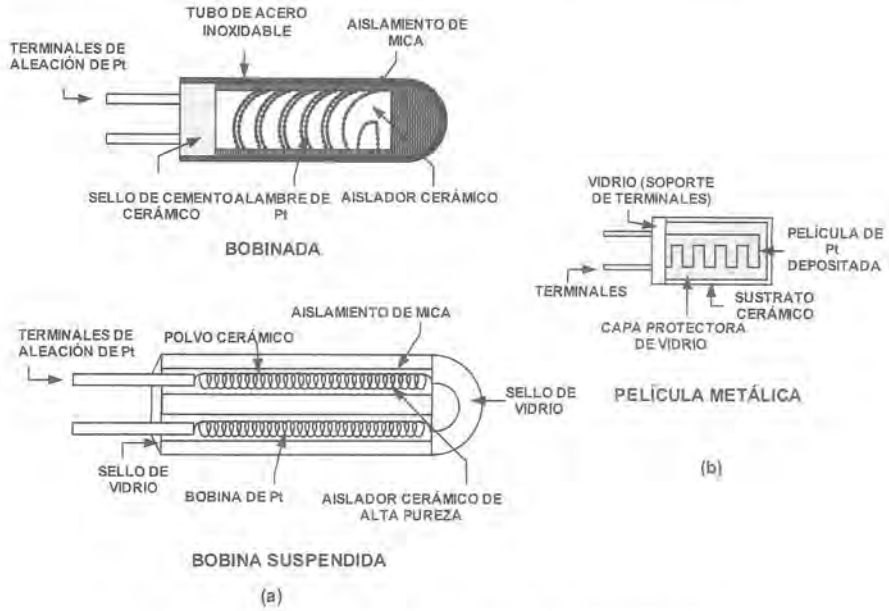


Figura 6.4. Tipos de RTDs según su construcción.

b) **Tamaño:** muy vinculado a la velocidad de respuesta, podemos encontrar desde RTDs en miniatura utilizadas donde se requiera medir la temperatura en un punto del medio y que, a la vez, la constante de tiempo sea muy pequeña, hasta RTDs de varios centímetros de longitud utilizadas en aplicaciones muy especiales para medir la temperatura promedio de una región del medio.

c) **Características del medio:** vinculado a su tecnología de fabricación, se diferencian las de inmersión en fluidos (líquidos y gases) de las de contacto superficial. Las de inmersión suelen ser bobinadas, protegidas con una envolvente metálica (generalmente acero inoxidable), mientras que las de contacto superficial generalmente son de película delgada montadas sobre sustratos planos o flexibles para adaptarse a la forma de las superficies.

d) **Tipo de metal utilizado:** las más populares son las de platino, por su amplio margen de trabajo lineal, baja reactividad con la mayoría de los medios en las aplicaciones en las que el hilo se encuentra en contacto directo con este, muy elevada estabilidad temporal y pequeñas deformaciones en un amplio margen de temperatura; también se construyen de wolframio (para trabajo en muy altas temperaturas y pocas exigencias de precisión), de cobre (para bajas temperaturas y pocas exigencias de precisión) e incluso de balco[3] (para bajas temperaturas y aplicaciones de muy bajo coste).

[3] Aleación de 70%Ni 30%Fe para mejorar la linealidad.

e) ***Valor de R_0:*** *también se pueden clasificar por el valor de la resistencia a cero grados centígrados,* siendo los valores más utilizados 100, 500 y 1000 Ω. La tabla 6.2 presenta esta información para las RTDs de los materiales más utilizados, incluyendo también el margen de temperaturas de trabajo.

Metal	Margen útil de temperatura (°C)	Valores de R_0 (Ω)
Platino	-260 ÷ 900	25, 100, 400, 500, 1000 y 2000
Wolframio	-100 ÷ 1200	10, 50, 100, 1000 y 2000
Níquel	-200 ÷ 430	120, 1000
Cobre	-200 ÷ 260	10, 100, 1000
Balco	-100 ÷ 230	100, 1000 y 2000

Tabla 6.2. Margen útil de temperatura y valores de R_0 para las RTDs comerciales [6.2], [6.3].

Este sensor es uno de los que más normativa internacional reúne para garantizar su sustitución en las aplicaciones. Aunque el platino sea un metal precioso y de alto coste, las RTDs construidas con este metal son las más difundidas por su amplio margen de trabajo, linealidad y estabilidad.

f) **Precisión:** existen normas que clasifican a las RTDs en función de su tolerancia. Por ejemplo, la norma UNE-EN60751 precisa dos clases de tolerancia para las RTDs de platino: clases A y B. En las de clase A, el error en la temperatura en °C se encuentra acotado en 0,15 ± 0,002T y en las de clase B en 0,3 ± 0,002T, expresado T en ambos casos en K [6.5].

6.3 El autocalentamiento

Para utilizar una RTD en un sistema de medida de temperatura, generalmente hay que hacerle pasar una corriente eléctrica. Esta corriente produce una disipación de energía en la misma que puede hacer que su temperatura t sea más alta que la del medio (t_a) en que se encuentra. El incremento de temperatura Δt que sufre la RTD debido al autocalentamiento puede evaluarse por:

$$\Delta t = t - t_a = R_\Theta \cdot I^2 \cdot R_t \ [°C] \tag{6.5}$$

donde: R_Θ (°C/W) es la resistencia térmica de la RTD con el medio, I es el valor efectivo de la corriente que circula por la RTD, R_t es la resistencia de la RTD a la temperatura t y t_a es la temperatura del medio.

Debe cuidarse que este error sea el mínimo permisible en el peor de los casos (R_t máxima) si se excita la RTD con una fuente de corriente constante.

EJERCICIO 6.1

Una RTD de platino se utiliza para la medición de la temperatura del aceite utilizado en un circuito de refrigeración. Se sabe que el margen de temperatura del aceite es 28 ÷ 87°C. Calcule la $I_{máx}$ que puede hacerse circular por la RTD si se desea que el error por autocalentamiento sea inferior a 0,1°C.

La Pt100 tiene $\alpha = 0{,}00385\ K^{-1}$ y $R_\Theta = 12{,}5°C/W$ en el aceite.

SOLUCIÓN:

Para la máxima temperatura de trabajo, $R_t = 100(1 + 0{,}00385 \cdot 87) = 133{,}5\ \Omega$.

Entonces de (6.5):

$$I_{máx.} = \sqrt{\frac{0,1}{(12,5)\cdot(133,5)}} = 7,74 \text{ mA}$$

6.4 Aplicaciones

Las RTDs son los sensores de temperatura más utilizados hoy en día tanto en aplicaciones industriales como domésticas, en los controladores de temperatura de los equipos de frío y calor, en los coches, en los hornos domésticos, climatizadores y cámaras climatizadas de precisión. Esta gran difusión de su aplicación se debe a su gran estabilidad en el tiempo, muy buena precisión (del orden de ±0,2°C para las de platino de uso industrial en el margen de –200 ÷ 500°C) y hasta ±0,0001°C para aplicaciones especiales de laboratorio que hace que se use como patrón de medida para las temperaturas comprendidas entre –183°C y 630,8°C. Gracias a su gran variedad de encapsulados, así como a su capacidad de permanecer inertes en aplicaciones en que el hilo conductor tenga que interactuar con el medio directamente para una respuesta rápida, cada día aumenta su campo de aplicación.

6.5 Acondicionamiento de la señal

La RTD ofrece un cambio de resistencia eléctrica como respuesta al cambio de la temperatura que es la que se pretende medir. Medir la temperatura implica, entonces, medir la resistencia de la RTD y, mediante su curva o tabla de calibración, evaluarla. Pero este método no ofrece una señal eléctrica proporcional a la temperatura objeto de medición que pueda ser registrada o medida fácilmente. Si se hace circular una corriente I conocida (con la limitación que impone el autocalentamiento) por la RTD y se mide la tensión a través de la misma, esta señal eléctrica será directamente proporcional a la temperatura.

En la práctica, se emplean varios sistemas acondicionadores para evaluar las variaciones de R_t. Entre los más generalizados se encuentran los puentes de resistencias alimentados con fuentes de tensión o de corriente continua y las conexiones directas de la RTD con sistemas amplificadores de instrumentación a 2, 3 y 4 hilos.

6.5.1 PUENTE DE WHEATSTONE ALIMENTADO CON TENSIÓN

El propósito del acondicionador de una RTD es lograr una señal de tensión proporcional a la temperatura a medir. Un circuito que realiza esta función es el puente de Wheatstone (véase Anexo 1 en el CD adjunto) alimentado en tensión de la figura 6.5 que proporciona una tensión diferencial dependiente de las variaciones de las resistencias de sus ramas.

Si consideramos el modelo lineal de la RTD, donde $R_t = R_0(1+\alpha t)$, se puede demostrar que el valor de la tensión de salida V_s en el circuito será:

$$V_s = V_b - V_a = V \cdot \frac{R_0(1+\alpha t)}{R_0(1+\alpha t)+R_2} - V \cdot \frac{R_3}{R_1+R_3} \qquad (6.6)$$

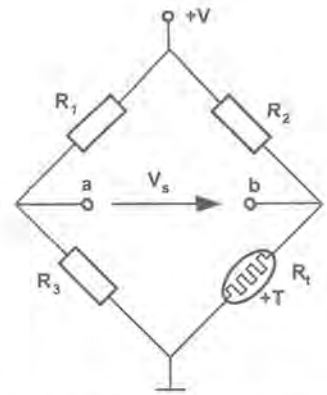


Figura 6.5. Puente de Wheatstone. Medida por deflexión.

Si deseáramos obtener tensión nula en la salida del puente ($V_s = 0V$) para $t = 0°C$, el signo de V_s se corresponderá con el de la temperatura a medir; entonces, en el puente se cumple que:

$$\frac{R_0}{R_0 + R_2} = \frac{R_3}{R_1 + R_3} \tag{6.7}$$

En los puentes es común colocar resistencias iguales en las ramas superiores y de un valor r veces mayor que la resistencia R_0 de la RTD ($R_1 = R_2 = R = rR_0$), así como seleccionar $R_3 = R_0$ con lo que el circuito mantendrá muy buena simetría si las variaciones de R_t son pequeñas y, a la vez, cumplirá con la ecuación (6.7). Si en la ecuación (6.6) hacemos estas sustituciones obtenemos:

$$V_s = V \cdot \frac{r \cdot \alpha t}{(r+1)\cdot(r+1+\alpha t)} \tag{6.8}$$

donde $r = R/R_0$ es la razón característica de resistencias del puente.

Con el puente de Wheatstone hemos obtenido una señal eléctrica V_s dependiente de la temperatura t aunque también de la igualdad entre las resistencias $R_1 = R_2 = rR_0$; $R_3 = R_0$, y del valor y estabilidad de la tensión de alimentación del puente V. Si se utilizan en el circuito resistencias de gran calidad, así como fuentes de alimentación muy estables o referencias de tensión (Anexo 1 en el CD adjunto), el puente no introduce errores apreciables en la medida. Sin embargo, la dependencia de la tensión de salida con la variable a medir (temperatura de la RTD) no es lineal, siendo esta la principal desventaja del puente alimentado con tensión.

6.5.1.1 Linealidad

Como puede observarse de la ecuación (6.8), la dependencia de la tensión de salida del acondicionador V_s con la temperatura t no es lineal, y lograr la linealidad del sistema sensor-acondicionador es una de las exigencias del acondicionador analógico en cualquier sistema de medida.

Pero en el caso en que la razón característica de resistencias del puente sea mucho mayor que el producto $\alpha \cdot t$ (para la mayoría de las aplicaciones con RTD comerciales $|\alpha \cdot t|$ no suele ser mayor de 0,5), entonces la expresión (6.8) puede aproximarse al modelo lineal:

$$V_{sL} = V \cdot \frac{r \cdot \alpha \cdot t}{(r+1)^2} \tag{6.9}$$

El uso del modelo anterior para el acondicionador implica que se comete un error de aproximación o error de *linealidad* que es importante evaluar.

La expresión del error de linealidad relativo al valor real (ε_{LV}) para el puente excitado con tensión V y para cualquier valor de αt será:

$$\varepsilon_{LV}[\%] = \frac{V_s - V_{sL}}{V_s} \cdot 100\% = -\frac{\alpha \cdot t}{r+1} \cdot 100\% \tag{6.10}$$

El signo de ε_{LV} indicará si la aproximación lineal realizada tiene un error de linealidad por exceso (+) o por defecto (−). El error de linealidad disminuye al aumentar la razón característica de resistencias del puente r, pero al propio tiempo, la tensión de salida V_{sL} es menor, por lo que la sensibilidad del sistema de acondicionamiento es menor.

6.5.1.2 Sensibilidad

La sensibilidad del sistema RTD-Puente es la razón de cambio de la salida del puente V_s respecto al cambio de la temperatura de la RTD y es una medida directa de la ganancia del sistema sensor-acondicionador.

Si en (6.8) obtenemos la derivada de V_s respecto a la temperatura t, se obtiene que la sensibilidad del sistema S_V será:

$$S_V = \frac{dV_s}{dt} = V \cdot \alpha \cdot \frac{r}{(r+1+\alpha \cdot t)^2} \ [V/°C] \tag{6.11}$$

La sensibilidad es directamente proporcional a la tensión de alimentación del puente pero, en la misma medida que la razón característica de resistencias del puente sea mayor con el propósito de lograr mejor linealidad, la sensibilidad se reduce. Además, es dependiente de la señal a medir y varía en forma no lineal con ella.

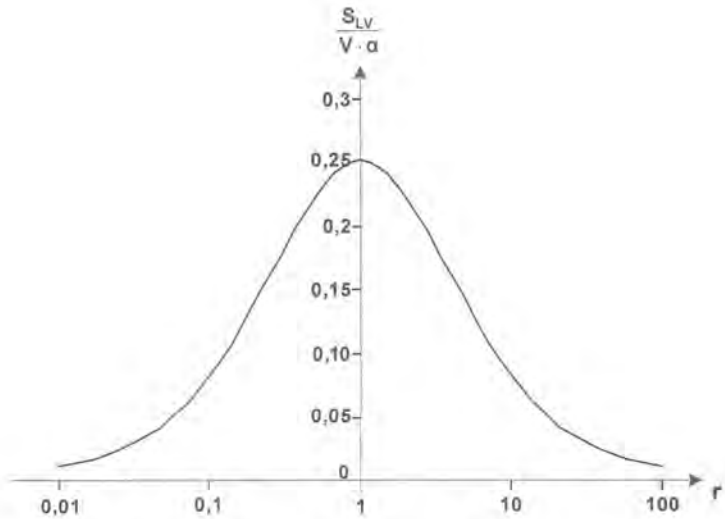


Figura 6.6. Variación de la sensibilidad normalizada del sistema con la razón característica de resistencias del puente alimentado con tensión.

Si se puede considerar el sistema lineal, derivando respecto a la temperatura la ecuación (6.9), podemos obtener:

$$S_{LV} = \frac{dV_{sL}}{dt} = V \cdot \alpha \cdot \frac{r}{(r+1)^2} \ [V/^oC] \tag{6.12}$$

La sensibilidad es **independiente de la variable a medir**, sólo depende de la razón característica de resistencias del puente. La figura 6.6 presenta la dependencia de la sensibilidad S_{LV} normalizada al producto $V \cdot \alpha$ con la razón característica de resistencias del puente r.

Podemos observar que la sensibilidad normalizada al producto $V\alpha$ es máxima para r = 1 e igual a 0,25, pero sabemos que para r = 1 en el puente realmente se produce un importante error de linealidad. Por lo que en el diseño de este acondicionador surge un compromiso entre linealidad y sensibilidad.

EJERCICIO 6.2

Para la RTD y el margen de medida del ejercicio 6.1, diseñe un puente de medida excitado con tensión de respuesta V_s lineal con t si se permite un error relativo de no-linealidad del 1%. Especifique la señal de tensión de salida y elabore el modelo del puente visto desde su salida.

SOLUCIÓN:

En el máximo valor de temperatura se encontrará el mayor error de linealidad. Esto se produce para $\alpha t = 0{,}00385 \cdot 87 = 0{,}33495$.

Si el error permitido máximo es de 1%, entonces,

$$r \geq \left| \frac{-\alpha \cdot t}{\varepsilon_{LV}} - 1 \right| = \left| \frac{-0{,}335}{0{,}01} - 1 \right| = 34{,}5$$

Escojamos r = 35. Así las cosas, $R_3 = 100\ \Omega$; $R_1 = R_2 = 3{,}5\ k\Omega$.

Si seleccionamos como $I_{máx} = 5mA$ por la RTD, entonces la tensión de excitación V será:

$$V < 0{,}005 \cdot (R_{t=87^oC} + 3500) = 0{,}005 \cdot (133{,}5 + 3500) = 18{,}17\ V$$

Escogemos V = 15V. La expresión de la curva de calibración lineal del sistema RTD-Puente será:

$$V_{SL} = V \cdot \frac{r \cdot \alpha \cdot t}{(r+1)^2} = 15 \cdot \frac{35 \cdot 0{,}00385}{36^2} \cdot t = 0{,}00156 \cdot t \quad \text{para} \quad 28^oC \leq t \leq 87^oC$$

La sensibilidad del sistema es de apenas 1,56 mV/°C. La asociación salida del puente-temperatura será:

Para $t_1 = 28°C$;

$$V_{SL1} = V \cdot \frac{r \cdot \alpha \cdot t_1}{(r+1)^2} = 15 \cdot \frac{35 \cdot 0{,}00385 \cdot 28}{36^2} = 43{,}7\ mV$$

Para $t_2 = 87°C$;

$$V_{SL2} = V \cdot \frac{r \cdot \alpha \cdot t_2}{(r+1)^2} = 15 \cdot \frac{35 \cdot 0{,}00385 \cdot 87}{36^2} = 135{,}69\ mV$$

La tensión de modo común será:

$$V_{CM} = V_a = V \cdot \frac{R_3}{R_1 + R_3} = 15 \cdot \frac{100}{3600} = 0{,}417 \text{ V}$$

Esta tensión es independientemente del cambio de la resistencia en la RTD para el puente excitado con tensión.

Teniendo en cuenta que $R_A = rR_0^2/(r+1)R_0 = r.R_0/(r+1)$ y $R_B = rR_0(R_0(1+\alpha t))/(rR_0+R_0(1+\alpha t)) = rR_0(1+\alpha t)/(r+1+\alpha t)$, las resistencias equivalentes del puente vistas desde los puntos a y b (figura 6.5) son $R_A = 35 \cdot 100/36 = 97{,}222\ \Omega$ y $R_B = 35 \cdot 100(1+\alpha t)/(36+\alpha t)$.

Para t = 0 °C, $R_B = R_A = 97{,}222\ \Omega$, que corresponde a $V_s = 0$ V. Esta condición no se encuentra dentro del margen de medida del problema.

Para $\alpha t = 28°C \cdot 0{,}000385°C^{-1} = 0{,}01078$, luego $R_B = 98{,}24\ \Omega$
Para $\alpha t = 87°C \cdot 0{,}000385°C^{-1} = 0{,}33495$, luego $R_B = 128{,}59\ \Omega$

El modelo para el sistema como fuente de señal de tensión proporcional a la temperatura t se presenta en la figura 6.7.

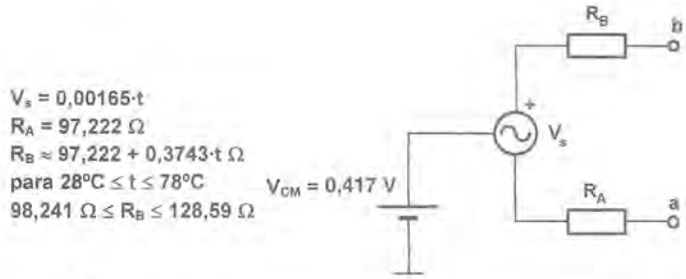


Figura 6.7. Modelo del sistema RTD-puente como fuente de señal.

6.5.1.3 Conexión remota del puente

Es muy frecuente que el puente no pueda estar físicamente ubicado cerca de la RTD. Por ello se emplean cables conductores más o menos largos que poseen una resistencia R_W, a menudo apreciable en comparación con la resistencia de la RTD y que sus cambios con la temperatura pueden afectar seriamente a la medida. En caso de utilizarse conductores de aleaciones especiales de muy pequeño coeficiente térmico de variación de resistencia tales como constantán o manganina para evitar los errores por cambio en la temperatura de los mismos, sus resistencias suelen ser muy altas. Si, por el contrario, se utilizan cables de cobre de muy baja resistencia, los cables pueden incorporar errores por su alto coeficiente térmico. Una solución la ofrece el "método Siemens" de conexión del puente o conexión a tres hilos, ilustrado en la figura 6.8.

Si los hilos conectados al puente en los puntos c y d son idénticos (longitud, material, calibre de la sección transversal), sus resistencias R_W son iguales. Considerando que en la rama derecha superior la resistencia es rR_0+R_W y que en la rama inferior es R_t+R_W, la expresión (6.8) queda ahora como:

$$V_s = V \cdot \frac{r \cdot \alpha \cdot t + (r-1) \cdot \frac{R_W}{R_0}}{(r+1) \cdot \left(r+1+\alpha \cdot t+2\frac{R_W}{R_0} \right)} \qquad (6.13)$$

El error relativo en la medida de la tensión V_s provocado por el valor de la resistencia de los hilos es:

$$\varsigma_r = \frac{(r-1-\alpha\cdot t)\cdot\frac{R_W}{R_0}}{(r+1+\alpha\cdot t)\cdot(r+1+\alpha\cdot t+2\cdot\frac{R_W}{R_0})} \qquad (6.14)$$

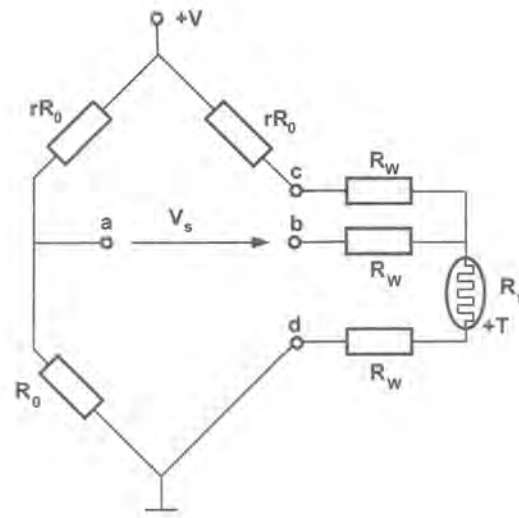


Figura 6.8. Conexión del puente a tres hilos.

Si para el sistema se cumple que la razón característica de resistencias del puente r es muy grande con respecto a los valores máximos que puede tomar αt en la aplicación, entonces (6.14) queda independiente de la señal a medir y puede simplificarse, quedando como:

$$\varsigma_r = \frac{(r-1)\cdot\frac{R_W}{R_0}}{(r+1)\cdot(r+1+2\cdot\frac{R_W}{R_0})} \qquad (6.15)$$

La figura 6.9 presenta la expresión gráfica del error ς_r con la relación (R_W/R_0) para la conexión a 3 hilos, utilizando como parámetro la razón característica de resistencias del puente r.

Puede observarse que, en la medida en que la razón característica de resistencias del puente es mayor, ς_r es menor. En el trabajo remoto de este acondicionador es aconsejable que para aplicaciones con conductores muy largos se seleccione, de ser posible, RTDs de alto valor de R_0 (1000 Ω, por ejemplo). Hemos supuesto que las resistencias de los hilos conectados a los puntos c y d tienen que ser iguales en todo momento, aunque esto no se logra en muchas aplicaciones debido, entre otros factores, a los efectos que produce el calentamiento desigual.

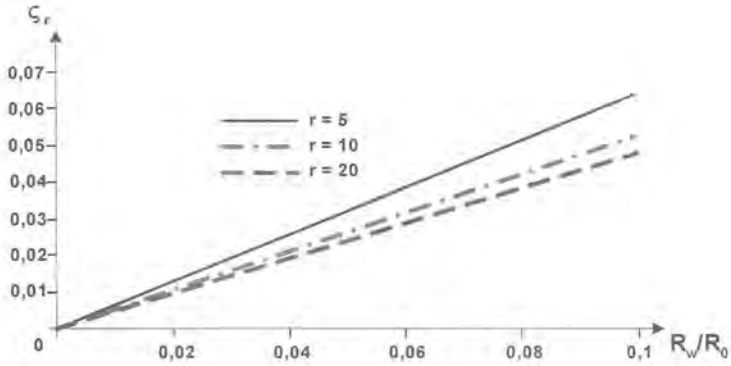


Figura 6.9. Error relativo de la medida de V_s provocado por la resistencia de los hilos.

Para reducir el error introducido por las resistencias de los hilos en medidas de buena precisión se suele utilizar la denominada conexión a 4 hilos. Con los hilos conectados como en la figura 6.10a se realiza una primera medida. En los extremos de conexión del puente se cambia la posición de los hilos, como muestra la figura 6.10b. En esta nueva posición se realiza otra medida. El resultado a ofrecer es el promedio de estas medidas, con lo que cualquier diferencia entre los valores de las resistencias R_W de los hilos a y d queda compensada.

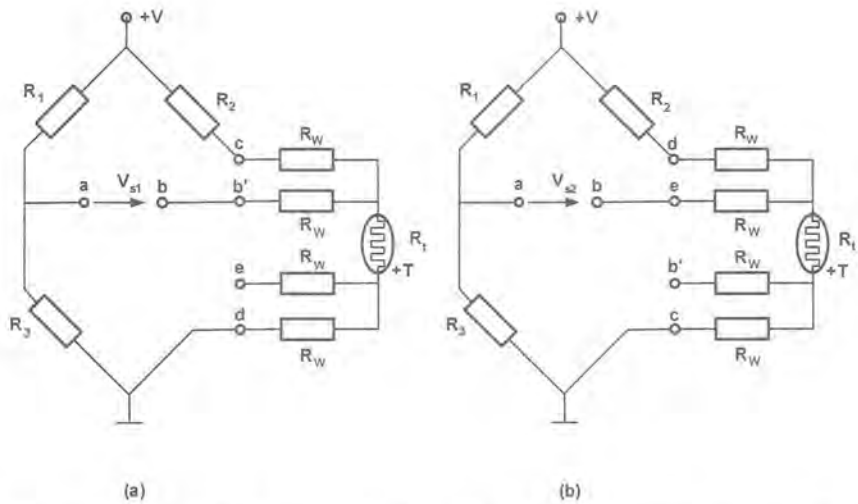


Figura 6.10. Conexión a cuatro hilos para el puente de deflexión, a) primera medida, b) segunda medida.

EJERCICIO 6.3

En un sistema de medida con una Pt100 con $\alpha = 0{,}00385K^{-1}$ se utiliza un puente que se encuentra a 150 m de la RTD. Si se utiliza una conexión a tres hilos entre el puente y la RTD y el cable es de cobre, con un área de la sección transversal $A = 1\ mm^2$ y una resistividad por unidad de longitud de $\rho = 0{,}018\ \Omega \cdot mm^2 \cdot m^{-1}$, ¿qué error se cometerá en la medición suponiendo que el valor de la resistencia de los cables no cambiará al cambiar la temperatura en la RTD y en el proceso? La razón característica de resistencias del puente es de 25.

SOLUCIÓN:

Según (6.1) la resistencia de los hilos con este cable será: $R_W = 0{,}018 \cdot 150/1 = 2{,}7\ \Omega$. Y el error relativo de la medida con estos cables en conexión a tres hilos con el puente será (de 6.15):

$$\varepsilon_r = ((25 - 1)\cdot 2{,}7/100)/(25 + 1)(25 + 1 + 2\cdot 2{,}7/100) = 0{,}000956 = 0{,}0956\%$$

Obsérvese el efecto sobre el error de medición que tiene la selección de los cables. Por ello en una conexión remota deben utilizarse los cables de la mayor sección posible.

6.5.2 PUENTE DE WHEATSTONE ALIMENTADO POR CORRIENTE

Si al puente de la figura 6.5 con $R_1 = R_2 = r \cdot R_0$ y $R_3 = R_0$ lo alimentamos con una fuente de corriente de valor $I = V/R_{eq}$, donde R_{eq} es la resistencia equivalente conectada a la fuente de tensión V que puede expresarse como la resultante paralelo de las resistencias de las ramas, a saber

$$R_{eq} = \frac{R_0 \cdot (r+1) \cdot (r+1+\alpha \cdot t)}{2r+2+\alpha \cdot t}$$

entonces la expresión (6.8) puede transformarse en:

$$V_s = I \cdot R_0 \cdot \frac{r \cdot \alpha \cdot t}{2 \cdot (r+1) + \alpha \cdot t} \tag{6.16}$$

resultado parecido al obtenido cuando se excita el puente con una fuente de tensión, es decir, obtenemos una señal de salida dependiente de la variable a medir y de la razón característica de resistencias del puente. Como puede observarse de la ecuación (6.16), la dependencia de la tensión de salida del acondicionador V_s con la temperatura t tampoco es lineal, como en el caso de excitar el puente con tensión.

Para el caso en que la razón característica de resistencias del puente sea mucho mayor que el producto αt (para la mayoría de las aplicaciones con RTD comerciales $|\alpha t|$ no suele ser mayor de 0,5), la expresión (6.16) puede aproximarse al modelo lineal:

$$V_{s_L} = I \cdot R_0 \cdot \frac{r}{2 \cdot (r+1)} \cdot \alpha \cdot t \tag{6.17}$$

De esta forma, la salida del puente queda linealmente dependiente de la temperatura y, si el valor de la razón característica de resistencias del puente es grande ($r >> 1$), en buena medida, independiente del mismo.

6.5.2.1 Linealidad

Cuando se alimenta el puente en corriente y se cumplen las condiciones que hacen posible la aplicación del modelo lineal en este, **ofrece un mejor comportamiento que cuando se excita el puente en tensión**; esto puede verse mejor evaluando el error de linealidad relativo:

$$\varepsilon_{LI}[\%] = \frac{V_s - V_{s_L}}{V_s} \cdot 100\% = -\frac{\alpha \cdot t}{2 \cdot (r+1)} \cdot 100\% \tag{6.18}$$

Si se compara (6.18) con (6.10) puede verse fácilmente que **el error de linealidad es la mitad cuando se excita en corriente que cuando se excita en tensión**. De esto se desprende que, en aplicaciones en las que se requiera gran linealidad, es aconsejable utilizar el puente excitado en fuente de corriente.

6.5.2.2 Sensibilidad

La sensibilidad del sistema se obtiene, al igual que cuando se excita en tensión, derivando respecto a la temperatura la expresión de la tensión de salida V_s (6.16), obteniendo:

$$S_I = \frac{dV_s}{dt} = 2IR_0\alpha \cdot \frac{r \cdot (r+1)}{(2r+2+\alpha \cdot t)^2} \tag{6.19}$$

Vemos de (6.19) que la sensibilidad del sistema depende de forma inversa y no lineal de la temperatura a medir.

Si se cumplen las condiciones para poder considerar el sistema como lineal, entonces la sensibilidad del sistema se obtiene al derivar (6.17) respecto a la temperatura t, resultando:

$$S_{LI} = \frac{dV_{sL}}{dt} = IR_0\alpha \cdot \frac{r}{2(r+1)} \tag{6.20}$$

La sensibilidad del sistema de medida alimentando el puente en corriente queda independiente de la temperatura a medir.

La figura 6.11 presenta la variación de la sensibilidad linealizada —expresión (6.20)— normalizada al producto $IR_0\alpha$ en función de la magnitud de la razón característica de resistencias del puente r.

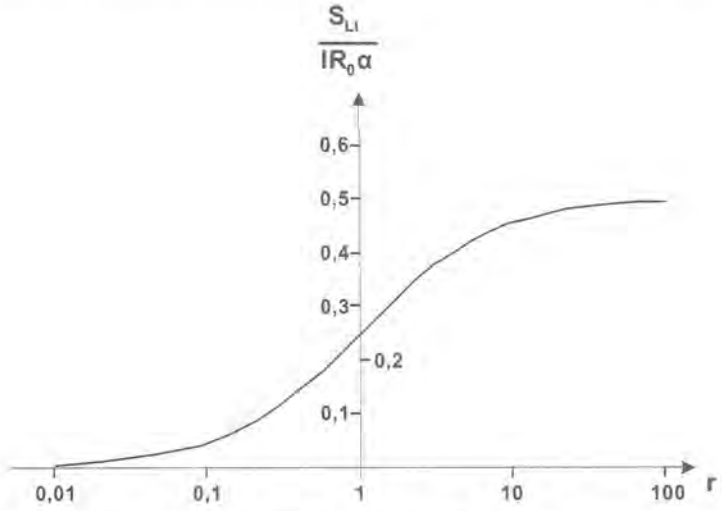


Figura 6.11. Variación de la sensibilidad normalizada con la razón característica de resistencias del puente alimentado en corriente.

Obsérvese que, a diferencia del puente alimentado en tensión, ahora la sensibilidad del sistema linealizado aumenta al aumentar la razón característica de resistencias del puente, teniendo 0,5 como valor de tendencia para la sensibilidad normalizada cuando r es muy grande.

EJERCICIO 6.4

Para la RTD y el margen de medida del ejercicio 6.1 diseñe un puente de medida excitado con corriente que ofrezca salida de tensión V_s lineal con la temperatura t si se permite un error relativo de no-linealidad del 1%. Especifique la señal de tensión de salida y elabore el modelo del puente visto desde su salida.

SOLUCIÓN:

El mayor error de linealidad se encontrará en el máximo valor de temperatura. Esto se produce para $\alpha t = 0{,}00385 \cdot 87 = 0{,}33495$.

Si el error máximo permitido es de 1%, entonces

$$r \geq \left| \frac{-\alpha \cdot t}{2 \cdot \varepsilon_{LI}} - 1 \right| = \left| \frac{-0{,}335}{0{,}02} - 1 \right| = 17{,}75$$

Escojamos r = 18. Así las cosas, $R_3 = 100\ \Omega$; $R_1 = R_2 = 1{,}8\text{k}\ \Omega$.

Si seleccionamos como $I_{2máx} = 5$ mA por la RTD, esto ocurrirá para la condición de R_{tmin}, correspondiente a t = 28°C. El valor de R_{tmin} será: $\alpha t_{min} = 28°C \cdot 0{,}00385°C^{-1} = 0{,}1078$ y $R_{tmin} = 110{,}8\ \Omega$.

La tensión a través de $(R_1 + R_{tmin})$ será: $V = 5 \cdot 10^{-3} \cdot (1800 + 98{,}241) = 9{,}4912$ V y la corriente por la rama $(R_1 + R_0)$ será: $I_{1min} = 9{,}4912/1900 = 4{,}995$ mA.

Escogemos $I_{1min} = 5$ mA. De esta forma, I = 10 mA.

La sensibilidad lineal del sistema de medida será:

$$S_{LI} = \frac{dV_{sL}}{dt} = IR_0\alpha \cdot \frac{r}{2r+2} = (0{,}01) \cdot 100 \cdot (0{,}00385) \cdot 18/38 = 0{,}001824\ \text{V}/°\text{C} = 1{,}824\text{m V}/°\text{C}.$$

La expresión de la tensión de salida V_{sL} será:

$$V_{s_L} = I \cdot R_0 \cdot \alpha \cdot \frac{r}{2 \cdot (r+1)} \cdot t = (0{,}01 \cdot 100 \cdot 0{,}00385 \cdot 18/38) \cdot t = 0{,}001824t \quad \text{para} \quad 28°C \leq t \leq 87°C$$

La asociación salida del puente-temperatura será:

Para $t_1 = 28°C$:

$$V_{s_{L1}} = I.R_0 \cdot \alpha \frac{r \cdot t_1}{2(r+1)} = 1 \cdot 0{,}00385 \frac{18 \cdot 28}{38} = 51{,}06\ \text{mV}$$

Para $t_2 = 87°C$:

$$V_{s_{L2}} = I.R_0 \cdot \alpha \frac{r \cdot t_2}{2(r+1)} = 1 \cdot 0{,}00385 \frac{18 \cdot 87}{38} = 158{,}66\ \text{mV}$$

Ahora la tensión de modo común será variable con la señal, de manera tal que:

para $t = t_1 = 28°C$, $V_{Cmmin} = I_{1min} \cdot R_0 = 0{,}005 \cdot 100 = 0{,}5$ V.
para $t = t_2 = 87°C$, $V_{Cmmax} = I_{1máx} \cdot R_0 = 0{,}005037 \cdot 100 = 0{,}5037$ V.

Las resistencias equivalentes que se ven desde los puntos a y b hacia masa son:

$$R_A = R_0 = 100\ \Omega$$
$$R_B = R_t = 100(1+0{,}00385t)$$

El circuito equivalente del sistema puente-RTD es el mismo que el del ejemplo 6.2, sólo que con estos nuevos valores.

Si se comparan los resultados obtenidos excitando con corriente con los del ejemplo 6.2, podremos concluir que:

- La sensibilidad del sistema alimentando el puente con corriente es mayor, en consecuencia la salida V_s es mayor.
- La tensión de modo común se reduce a la mitad, aunque varía con la señal de proceso.

6.5.3 OTROS CIRCUITOS ACONDICIONADORES PARA RTDS

Existen otros circuitos de acondicionamiento para RTDs muy utilizados en la práctica que requieren del concurso de amplificadores operacionales. En su conjunción con los puentes, realizan la función de linealizadores de la característica transferencial de los mismos.

Si en la red de realimentación de un simple inversor (figura 6.12a) se coloca la RTD, la salida del inversor será:

$$V_s = -\frac{V \cdot R_0}{R_1} \cdot (1 + \alpha t)$$

V_s es lineal con t con un *offset* en t = 0°C. Existe una componente de la tensión V_s que no depende de la temperatura t y que podría introducir un error considerable en la medida si la tensión V cambia con el tiempo y la temperatura. Si en el lugar de una fuente de valor +V se coloca una referencia de tensión V_{ref} (Anexo 1 en el CD adjunto), entonces los resultados ganan en precisión.

Si en el brazo del puente de Wheatstone correspondiente a R_t se inserta un amplificador operacional (figura 6.12b), se configura un amplificador diferencial. En éste, las resistencias conectadas entre la entrada inversora-salida y entrada no inversora-masa tienen un "desequilibrio" y la señal de entrada +V se corresponde con una tensión de modo común, por lo que la salida V_s será:

$$V_s = -\frac{V \cdot R_0}{R_1 + R_0} \cdot \alpha t$$

y variará linealmente con la temperatura t y el *offset* se corrige aunque se requiere una alta estabilidad temporal y térmica en la fuente de tensión V y en las resistencias del circuito.

Un resultado idéntico puede obtenerse con el circuito de la figura 6.12c, y se demuestra que:

$$V_c = -\frac{V \cdot R_0}{R_1}(1 + \alpha t) \text{ y que } V_s = -\frac{V \cdot R_0}{R_1 + R_0} \cdot \alpha t$$

Si se excita la RTD directamente con referencias de corriente I_{ref} (Anexo 1 en el CD) y con el concurso de amplificadores de instrumentación, también pueden obtenerse resultados muy satisfactorios con corrientes muy pequeñas ($I_{ref} \leq 1$ mA).

La figura 6.13 presenta la conexión de una RTD con un amplificador de instrumentación mediante dos, tres y cuatro hilos. En la conexión a dos hilos (figura 6.13a) se tiene:

$$V_s = I \cdot (R_{W1} + R_{W2} + R_t) \cdot G \qquad (6.21)$$

donde G es la ganancia diferencial del amplificador de instrumentación. La tensión de salida V_s tiene dos componentes: una tensión que depende linealmente de la temperatura de la RTD y otra que depende de los valores de la resistencia de los hilos y los contactos, así como de sus cambios temporales y térmicos. En el caso de utilizar conductores muy gruesos y distancias no muy grandes, de manera tal que se cumpla que la resistencia menor de la RTD sea mucho mayor que el doble de la resistencia del hilo R_W, entonces la expresión anterior puede aproximarse a:

$$V_s = G \cdot I \cdot R_0 + G \cdot I \cdot R_0 \cdot \alpha \cdot t \qquad (6.22)$$

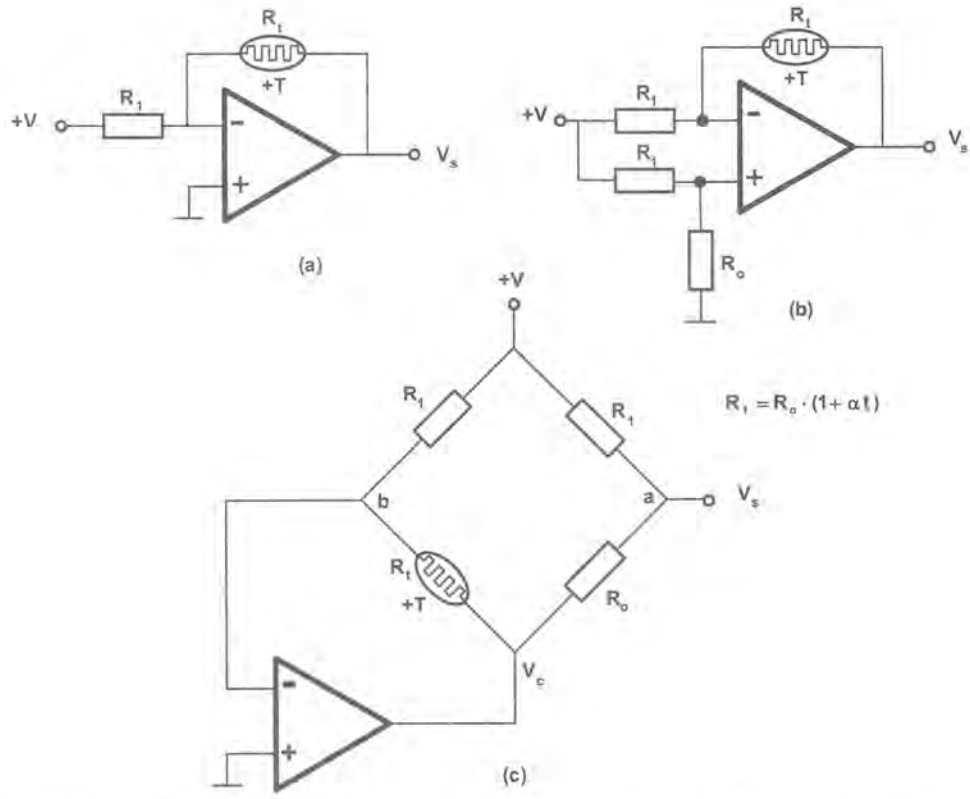


Figura 6.12. a) Amplificador inversor. b) Amplificador diferencial. c) Inversor con RTD en el bucle de realimentación.

En el mejor de los casos, es decir, cuando pueda aplicarse la aproximación (6.22), la tensión de salida es directamente proporcional a la temperatura de la RTD; no obstante, tiene un *offset* cuya estabilidad es dependiente de la estabilidad de la ganancia G y la corriente I, así como de la clase de la RTD.

En el caso de tres hilos (figura 6.13b), las fuentes de corriente I_1 e I_2 y los hilos conectados a las entradas del amplificador tienen que ser idénticos, respectivamente, para conseguir eliminar el efecto de las resistencias de los cables (con independencia de su longitud) y los contactos. Si se cumple que $R_{W1} = R_{W2}$ y que, además, $I_1 = I_2$, entonces

$$V_s = G \cdot (V_1 - V_2) = G \cdot I \cdot R_0 + G \cdot I \cdot R_0 \cdot \alpha \cdot t$$

resultando una expresión idéntica al caso de dos hilos pero en este caso realmente independiente de los valores de la resistencia de los hilos a costa de incorporar un hilo y una fuente de corriente más en el circuito.

En la conexión a cuatro hilos (figura 6.13c) se utiliza sólo una fuente de corriente I y los hilos no tienen por qué ser iguales ni de muy baja resistencia (siempre que no representen una resistencia de carga en serie *con R_i muy alta para la fuente de corriente que afecte su funcionamiento). Como idealmente no circula* corriente por los hilos conectados a las entradas del amplificador, la salida estará dada por:

$$V_s = G \cdot I \cdot R_0 + G \cdot I \cdot R_0 \cdot \alpha \cdot t$$

En la práctica, la corriente de polarización del amplificador incorpora un error proporcional al desequilibrio de la resistencia R_t de la RTD y los mejores resultados se obtienen aislando en el punto de medición la RTD de tierra física.

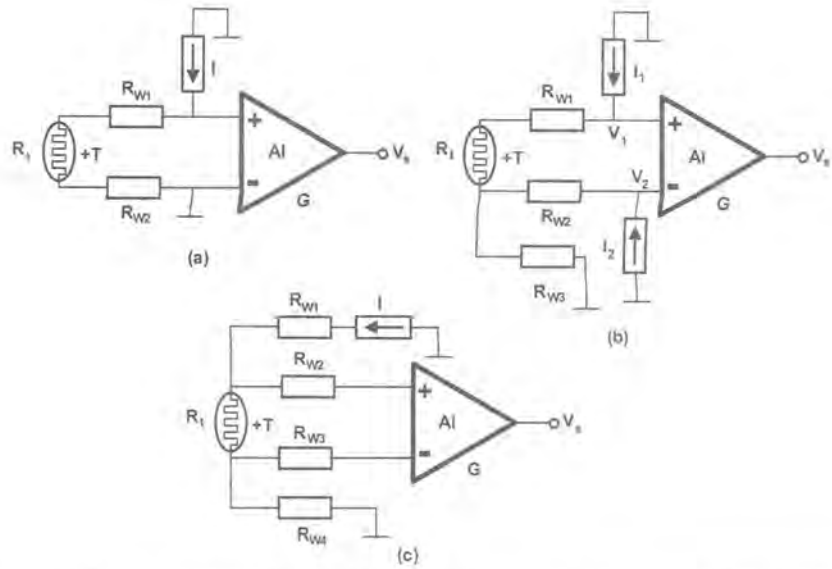


Figura 6.13. Conexión directa de RTDs alimentadas con referencias de corriente mediante conexiones de dos, tres y cuatro hilos.

Resumen

La RTD es un sensor de temperatura resistivo metálico de coeficiente térmico positivo muy utilizado en la práctica para la medición de la temperatura de medios y superficies, gracias a su gran variedad de formas constructivas. En particular, las de platino ofrecen las mejores características de linealidad y estabilidad temporal por lo que son las más utilizadas para la medición de temperaturas en el margen de –200 ÷ 800°C.

El circuito acondicionador por excelencia de RTDs es el puente de Wheatstone alimentado con fuentes de tensión y/o de corriente. El primer caso es el más utilizado en la práctica debido a la dificultad de lograr fuentes de corriente de algunos miliamperios con gran estabilidad temporal y térmica, aunque los resultados son mejores si se utilizan fuentes de corriente. Su uso conjunto con amplificadores de instrumentación que ofrezcan salidas de tensión para excitar puentes debe ser evaluado como una solución ventajosa.

Las RTDs pueden ser conectadas directamente a amplificadores de instrumentación en conexiones de 2, 3 y 4 hilos siendo la más eficiente la conexión a 4 hilos.

Problemas propuestos

6.1 Una Pt100 será utilizada para medir con alta precisión temperaturas comprendidas entre 0°C y 200°C. Considerando la ecuación (6.2) como modelo para la misma hasta su término cuadrático y

que los valores reales de R_t para 100°C y 200°C son 138,50 Ω y 175,83 Ω, respectivamente, calcule: a) los valores de α y β; b) si se considera lineal (β = 0), ¿qué error de linealidad se comete a 200°C?

6.2 Para el sistema del ejercicio 6.4 diseñe la fuente de corriente necesaria a partir de utilizar una referencia de tensión de 2,5 V y los esquemas de las figuras A1.14 o A1.15 del Anexo 1 en el CD.

6.3 Especifique la ganancia del amplificador de instrumentación para lograr una señal de salida en el margen de 0 ÷ 5 V si la temperatura medida por la RTD utilizada se puede encontrar entre 0°C y 80°C. Utilice una referencia de corriente tipo REF200 y conexión a 4 hilos. Datos de la RTD: Cu, $R_0 = 100\ \Omega$; $\alpha = 0{,}00421\ K^{-1}$.

7

Galgas extensométricas

7.1 Principio de funcionamiento
7.2 Tipos de galgas extensométricas
7.3 Utilización de las galgas extensométricas
7.4 Circuitos de medida
7.5 Utilización de los circuitos de medida
7.6 *Aplicaciones*

Las galgas extensométricas (*strain gauges*) son sensores resistivos muy utilizados en la medida de esfuerzos mecánicos en materiales en los que la resistencia efectiva entre sus extremos se modifica con el esfuerzo aplicado sobre la galga. Para comprender su funcionamiento, veremos en qué principios se basa el cambio de la resistencia con la fuerza aplicada.

7.1 Principio de funcionamiento

Supongamos un hilo de metal homogéneo de longitud l y diámetro d sujeto entre dos puntos según se muestra en la figura 7.1a.

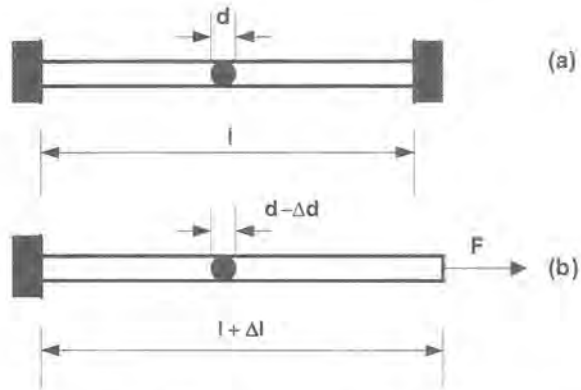


Figura 7.1. Deformación de un sólido sometido a tracción: (a) situación original; (b) modificación de la longitud y del diámetro con la fuerza de tracción aplicada.

La resistencia asociada al hilo de metal será proporcional a su longitud e inversamente proporcional a su sección, es decir:

$$R = \rho \frac{l}{\frac{\pi d^2}{4}} \tag{7.1}$$

donde R es la resistencia del hilo y ρ, su resistividad. Supongamos también que se aplica una fuerza F de tracción a uno de los extremos con lo que el material se deformará elásticamente (figura 7.1b). En principio, se modificarán las dimensiones produciéndose un alargamiento del hilo y una disminución de su sección que ocasionará un cambio en la resistencia efectiva del hilo (7.1), cambio que quedará recogido en la expresión:

$$\frac{\Delta R}{R} = \frac{\Delta \rho}{\rho} + \frac{\Delta l}{l} - 2\frac{\Delta d}{d} \tag{7.2}$$

que se obtiene tomando logaritmos neperianos en la expresión (7.1) y obteniendo la expresión incremental. Por otro lado, la ley de Poisson establece la deformación elástica de la sección de un cilindro de material cuando se somete a un esfuerzo de tracción o compresión en relación con la deformación longitudinal mediante un parámetro adimensional, el denominado coeficiente de Poisson[1]:

$$\upsilon = -\frac{\Delta d/d}{\Delta l/l} \tag{7.3}$$

Este coeficiente se suele expresar en microdeformaciones (ppm) de tal forma que una microdeformación es equivalente a 10^{-6} m/m = 1 μm/m.

Dividiendo todos los términos de la expresión (7.2) por Δl/l y teniendo en cuenta (7.3), se obtiene el denominado factor de galga, K:

$$K = \frac{\Delta R/R}{\Delta l/l} = 1 + 2\upsilon + \frac{\Delta \rho/\rho}{\Delta l/l} \tag{7.4}$$

Este factor determina la sensibilidad de la resistencia a los cambios en la longitud. Cuando sobre una galga se produce una deformación por unidad de longitud de valor Δl/l, el cambio en la resistencia puede deberse a:

- El propio cambio de longitud.
- El cambio originado en la sección.
- El cambio originado en la resistividad.

La expresión (7.4) tiene en cuenta estos cambios para obtener el factor de galga que, en la mayoría de los metales usados para realizar galgas extensométricas, toma valores entre 2 y 5.

Teniendo en cuenta la fuerza aplicada al hilo F, su sección $A = \pi d^2/4$ y el módulo de Young del material E, se tiene que las elongaciones por unidad de longitud Δl/l serán:

$$\varepsilon = \Delta l/l = \frac{F}{EA}$$

con lo que el cambio de resistencia resultará ser:

$$\Delta R = \frac{KR}{EA} F \tag{7.5}$$

[1] El valor del coeficiente de Poisson está comprendido entre 0 y 0,5 pero, en el caso particular de los metales, el margen es más estrecho, pudiendo considerarse que $0{,}24 \leq \upsilon \leq 0{,}4$.

Así, la sensibilidad de la galga crece con el factor de galga y el valor de la resistencia y decrece con el módulo de Young y con la sección. Para lograr mayores cambios en ΔR, será conveniente incrementar el valor de la resistencia sin incrementar la sección, es decir, tener mayores longitudes; por otro lado, si lo que interesa es medir en una zona muy concreta, el tamaño no puede ser muy grande con lo que se impone una solución en la que el hilo se disponga en zig-zag y se sitúe sobre una película que transmita lo mejor posible el estado tensional al propio hilo. En la figura 7.2 aparece una disposición de este tipo que explica cómo se construían las primeras galgas extensométricas.

Interesa que la galga mida esfuerzos en una sola dirección; con la disposición de la figura 7.2b mide esfuerzos en el eje vertical (y) mientras que los esfuerzos en el eje horizontal (x) no debieran afectar porque la galga se pliega como un acordeón. Sin embargo, en las zonas señaladas con círculos en la figura 7.2b la dirección del cable coincide con la del esfuerzo e introduce un error. Para reducirlo, se aumenta la sección A en esta zona —de acuerdo a (7.5)— con lo que se reduce la sensibilidad para el esfuerzo transversal en esos puntos.

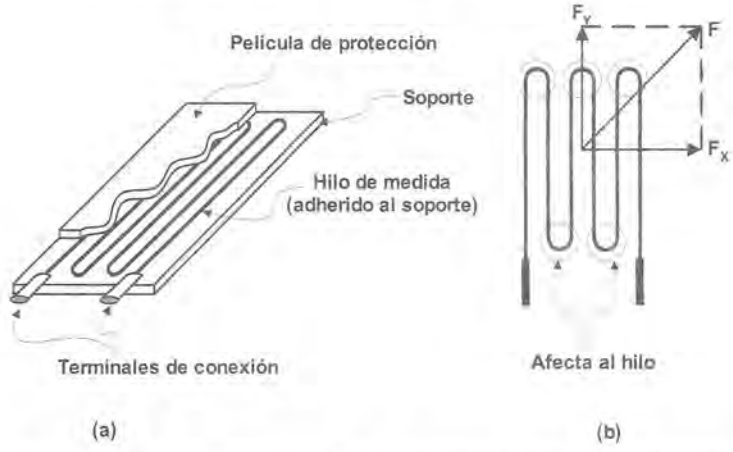


Figura 7.2. Galgas extensométricas de hilo: (a) construcción; (b) respuesta a la fuerza (los tramos de pista paralelos a F_x introducen un error).

Pero aparecen una serie de problemas asociados con este tipo de galga como son la respuesta ante estados tensionales no longitudinales que introducen un error en la medida (figura 7.2b), así como una pérdida de energía mecánica en la película intermedia y en la unión de esta con el propio hilo.

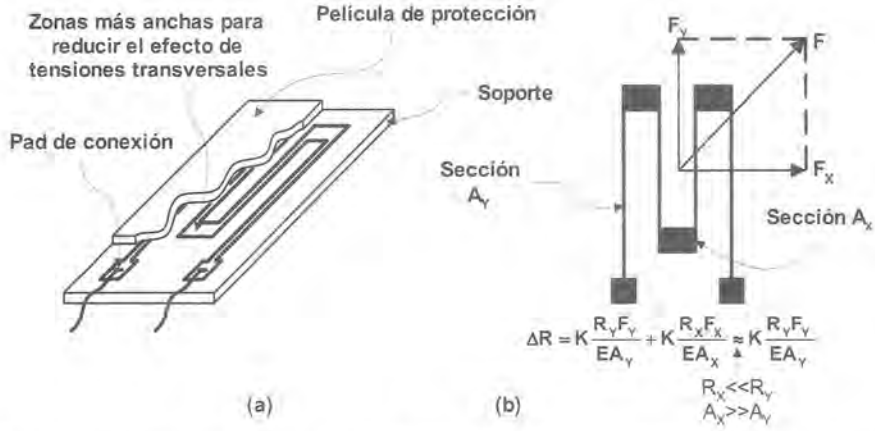


Figura 7.3. Galga extensométrica típica: (a) configuración física; (b) efecto de la fuerza sobre la galga.

Las galgas actuales se construyen con una película metálica depositada sobre el soporte (figura 7.3a) mediante un proceso de *sputtering* que reduce los efectos de los esfuerzos transversales mediante zonas de mucha más sección y que, por ello, contribuyen menos al error en la medida (figura 7.3b).

7.2 Tipos de galgas extensométricas

Las galgas que más se usan hoy en día son las galgas metálicas como las que se muestran en la figura 7.3 pero, además, se tienen otros tipos de galgas no metálicas como las de semiconductor; en este caso, el efecto del cambio de la resistencia con el estado tensional aplicado no se debe a las causas explicadas hasta este momento, sino a un *fenómeno piezorresistivo*.

El aspecto de las galgas de semiconductor es como el que se muestra en la figura 7.4 aunque en muchas ocasiones se incluyen varias galgas para formar un puente e, incluso, las etapas amplificadoras. Las galgas de semiconductor suelen tener factores de galga más elevados, entre 30 y 120, lo que las dota de mayor sensibilidad; sin embargo, el mercado sigue ampliamente dominado por las galgas metálicas y eso que el fenómeno piezorresistivo se conoce desde hace más de cincuenta años.

Figura 7.4. Galga extensométrica de semiconductor [7.2].

Los motivos de que exista esta situación de dominio de las galgas metálicas hay que buscarlos en las siguientes causas:

a) La deposición del metal sobre soportes flexibles es muy sencilla lo que proporciona láminas flexibles de fácil manejo y bajo coste. Además, el bajo coste de las láminas —en general, de poliamida— y la facilidad del proceso produce un amplio abanico de posibilidades que nos permite disponer de puentes completos, rosetas o cualquier otra topología.

b) Las galgas metálicas suelen tener *pads* de conexión de tamaño suficiente como para facilitar esa tarea, mientras que las dimensiones de los *pads* de las galgas semiconductoras son pequeños si se desea que su coste no se dispare. Además, exigen una soldadura compleja ya sea mediante resinas conductoras o mediante soldadura por ultrasonidos.

c) El coste final de las galgas semiconductoras es entre 10 y 20 veces superior al de las metálicas [7.1], diferencia de coste que se reparte entre los superiores costes del material, el bajo rendimiento

del proceso de fabricación, la baja repetibilidad, el coste de los sistemas de test y sólo una parte al menor tamaño del mercado.

d) Mayor sensibilidad térmica propia de un material semiconductor frente al comportamiento menos interferido que manifiestan los materiales metálicos.

No obstante, sí que se disponen algunos ejemplos de galgas piezorresistivas, sobre todo en aplicaciones concretas en que se integra el puente de medida, el amplificador y las compensaciones térmicas [7.1].

Las galgas metálicas utilizan diversos materiales para la medida y como soporte del metal en función de las prestaciones a obtener del sensor. En la tabla 7.1 se muestran algunas posibilidades y las características que se pueden esperar de ellas [7.7].

Materiales metálicos sensores		
Material	**Características**	**Aplicaciones**
Constantán	– Medidas estáticas o pseudo-estáticas – No se pueden usar en aplicaciones extremas – Selección compleja (pocos criterios) – Material más usado y muy barato – Autocompensación térmica sencilla	– Grandes elongaciones (estado plástico de deformación)
Isoelastic	– Gran relación S/N – Precisan un buen control de temperatura	– Medida de esfuerzos dinámicos – Medida de fatiga
Karma	– Autocompensación térmica sencilla – La soldadura de terminales es compleja	– Medida a temperaturas muy bajas – Medida con temperaturas variables o no controladas
Aleaciones de Pt	– Coste alto	– Medida a altas temperaturas
Materiales para el soporte		
Material	**Características**	**Aplicaciones**
Poliamida	– Es el soporte estándar – No soporta condiciones extremas de trabajo – Espesor habitual de 0,025 mm	– Medidas estáticas – Aplicaciones habituales
Epoxy	– Minimiza el error introducido por el soporte – Instalación delicada – Requiere mano de obra especializada	– Medidas precisas
Fibra de vidrio reforzada con epoxy	– Soporta temperaturas moderadas – Soporta muy bien el trabajo a fatiga	– Medidas cíclicas y de fatiga

Tabla 7.1. Materiales para galgas extensométricas. Características y aplicaciones.

Las galgas metálicas se pueden construir también mediante técnicas híbridas, depositando la pasta metálica sobre un sustrato de alúmina (Al_2O_3) y calentando el conjunto a unos 850°C con lo que se obtienen espesores entre 2 y 5 micras. Estas galgas tienen una sensibilidad algo superior a la que poseen las de película metálica pero bastante inferior a las de semiconductor. La soldadura y manejo son bastante cómodos pero poseen peores características térmicas [7.1].

La forma de medir el comportamiento de una galga frente a cambios en la temperatura es mediante dos parámetros: TCR (Coeficiente térmico de la resistividad) y TCGF (Coeficiente térmico del factor de galga); el primero mide la variación de la resistividad ρ en función de la temperatura, mientras que el segundo mide la variación de K respecto de T. En la tabla 7.2 se muestra un resumen de comparación de características entre los tres tipos de galgas que se han comentado.

Tipo	K	TCR ($\times 10^{-6}$/K)	TCGF ($\times 10^{-6}$/K)	Deriva Temporal
Metálica	2	10	100	Muy baja
Híbridas de capa fina	3 a 20	50	300	Baja
Semiconductor	30 a 120	1500	2000	Media

Tabla 7.2. Características comparadas de los tres tipos de galgas.

El parámetro de deriva temporal que se indica en la última columna de la tabla 7.2 establece las variaciones que sufre la galga extensométrica a lo largo de su vida útil.

El efecto del cambio de la temperatura sobre la resistencia efectiva de la galga introduce un error ya que ese cambio se puede interpretar como un cambio en el estado tensional; en estos casos, se suele hablar de deformación aparente y se suele medir en microdeformaciones por grado Kelvin (μ/K) tal y como se indica en la tabla 7.2.

7.3 Utilización de las galgas extensométricas

La correcta utilización de las galgas extensométricas, sean del tipo que sean, deben garantizar que lo que se mide sea lo que realmente se quiere medir. Para ello, hay que tener en cuenta lo siguiente:

a) Las galgas miden en una dirección exclusivamente, por lo que cuando se colocan en un determinado lugar, sólo son capaces de cambiar su resistencia en función de las componentes tensionales en la dirección de medida. Para ello, suelen incluir marcas que permiten su correcta alineación (figura 7.5).

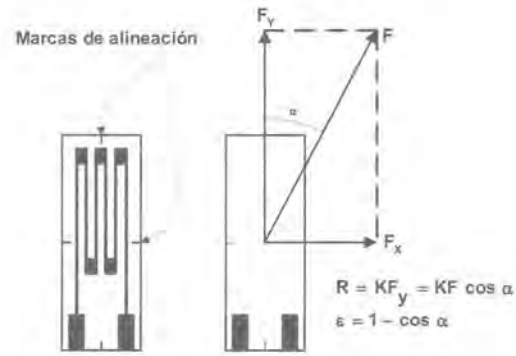


Figura 7.5. Utilización de las galgas extensométricas y descomposición de la fuerza aplicada según los dos ejes señalados por las marcas de alineación.

b) La medida de la tensión de la superficie se lleva a cabo en la zona que ocupa la galga y no será capaz más que de proporcionar un valor medio de toda la superficie involucrada por lo que no permite la detección de microestados tensionales.

c) Si se desconoce la dirección de la tensión a que estará sometida la galga o no se tiene garantía de que vaya a tener una dirección definida, se podrán utilizar dos galgas dispuestas en la misma zona ortogonalmente (figura 7.6) con lo que se tendrá una información sobre las dos componentes de la

tensión y, por ello, del estado tensional en esa zona. En el caso de la figura 7.6b, que suele ser muy normal, se supone que no hay cambios tensionales en toda la zona que ocupan las dos galgas.

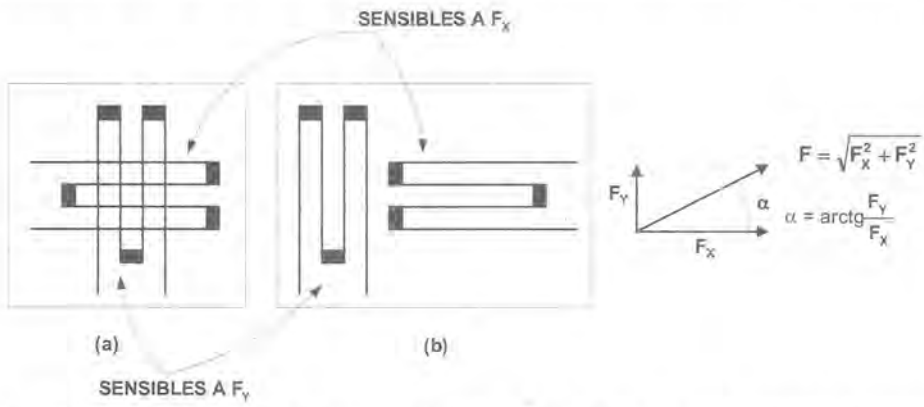


Figura 7.6. Utilización de dos galgas sobre el mismo soporte para la medida del vector de deformación definido por el módulo y por la dirección.

d) Si los estados tensionales no son planos, es decir, presentan una componente adicional, no será posible la determinación del estado ya que nunca se dispondrá de la tercera componente (la que resulta normal a la superficie en que se sitúa la galga).

e) La transmisión del esfuerzo de la superficie en que se va a medir hasta el metal debe realizarse sin pérdida de información, por lo que el sistema de sujeción debe garantizar la no absorción del esfuerzo. Para ello se emplean pegamentos de gran resistencia mecánica que permiten un óptimo funcionamiento. También se debe tener una cierta precaución en la "pasivación" de la superficie expuesta mediante una capa de material que la proteja frente a inclemencias ambientales[2]. En la figura 7.7 se muestran los pasos a realizar para la correcta utilización de las galgas extensométricas.

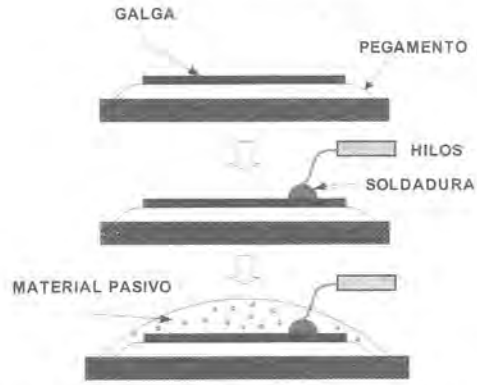


Figura 7.7. Pasos para la colocación de una galga sobre una superficie.

f) Los adhesivos utilizados dependen mucho de la aplicación. En la tabla 7.3 se indican algunas posibilidades en función del tipo de aplicación que se pretenda.

[2] Hay que recordar que las galgas extensiométricas se suelen usar en construcción con lo que quedan expuestas al ambiente, con el consiguiente peligro potencial de ser destruidas por cualquier golpe, por fricción, etc.

Tipo de adhesivo	Propiedades	Inconvenientes	Ejemplos
Cianocrilato	– Muy rápido (sólo precisa unos 10 minutos para poder ser usado) – Aplicar ligera presión 1 ó 2 minutos	– Conserva las propiedades durante sólo unos pocos meses	Eastman 910SL
Epoxy	– Gran resistencia – Permite medidas con grandes deformaciones – Se precisa una presión de 350 kPa en capas finas durante el proceso de curado	– Tiempo largo para el curado – A veces se precisa, aplicar 120°C durante algunas horas para completar el proceso	
Cerámico	– Alta temperatura	– Tiempo largo de curado	NBS-x-142 BLH-H
Nitrato de celulosa	– Galgas con soporte de papel – Aplicación sencilla	– Se precisa aplicar aire caliente (54°C) para el curado y la eliminación de disolventes – No puede trabajar en ambientes húmedos	Duco SR-4

Tabla 7.3. Adhesivos utilizados para el pegado de las galgas extensométricas.

Además de las disposiciones de galgas como las de la figura 7.6 (a y b) para medir estados tensionales hay más posibilidades que permiten medir esfuerzos radiales, axiales o en direcciones no ortogonales, alternativas que se muestran en la figura 7.8.

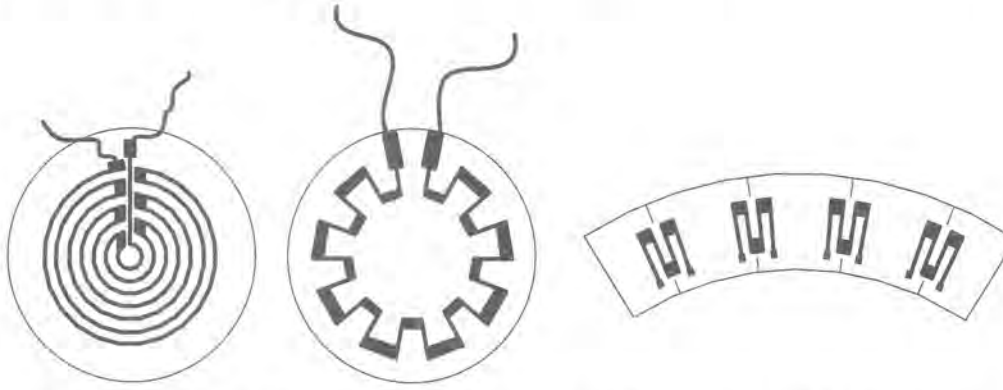


Figura 7.8. Otras configuraciones de galgas para medir esfuerzos radiales (izquierda), axiales (centro) o en diversas direcciones (derecha).

En cualquier caso, los fabricantes ofrecen una gama muy amplia de este tipo de sistemas que se podrán adaptar casi a cualquier necesidad.

7.4 Circuitos de medida

La medida de los estados tensionales mediante galgas extensométricas se lleva a cabo con circuitos similares a los que se usaron en otros sensores resistivos como las RTDs que se estudiaron en el capítulo anterior.

Como idea básica inicial, se usarán circuitos en puente aunque las especiales circunstancias de las galgas podrán modificar las topologías iniciales.

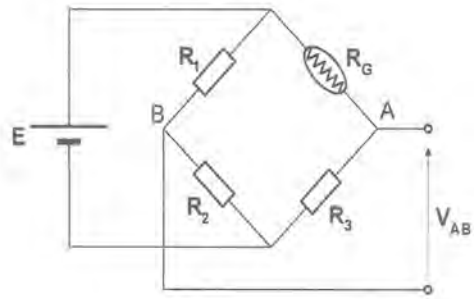


Figura 7.9. Puente básico de medida con una galga extensométrica.

En la figura 7.9 se muestra un circuito básico en puente que proporciona una tensión entre sus terminales de salida v_{AB}:

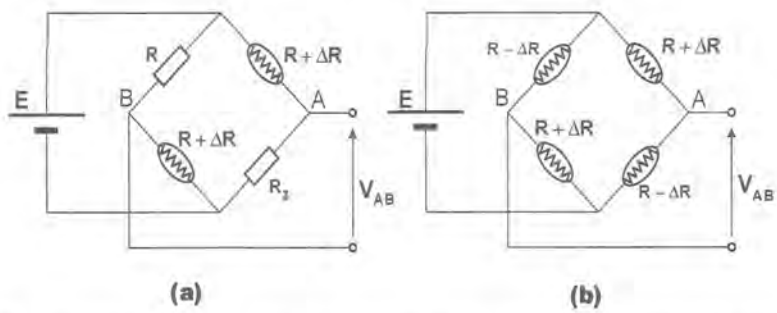
$$v_{AB} = E\left(\frac{R_3}{R_3 + R_G} - \frac{R_2}{R_1 + R_2}\right) \tag{7.6}$$

Si se hacen todas las resistencias iguales entre sí y de valor R, la expresión (7.6) se transforma en:

$$v_{AB} = E\left(\frac{R}{R + R_G} - \frac{1}{2}\right) \tag{7.7}$$

Considerando que la galga activa tiene una resistencia $R_G = R + \Delta R$, es decir, una resistencia propia R y otra debida a la elongación a que está sometida, ΔR. La expresión (7.7) se transforma en:

$$v_{AB} = -\frac{\Delta R}{4}\frac{1}{R + \frac{\Delta R}{2}}E \tag{7.8}$$



(a) **(b)**

Figura 7.10. Puentes de medida para incrementar la sensibilidad respecto del circuito básico con dos galgas activas (a) y con cuatro galgas activas (b).

Incluyendo la elongación ε, en la anterior expresión, se obtiene la relación entre ella y la tensión de salida del puente que incluye un término no lineal cuyo efecto puede despreciarse para pequeñas elongaciones:

$$v_{AB} = -\frac{K\varepsilon}{4}\frac{1}{1 + \frac{K\varepsilon}{2}}E \cong \frac{K\varepsilon}{4}E \tag{7.9}$$

Si se disponen dos galgas extensométricas trabajando con el mismo esfuerzo aplicado (muy próximas entre sí), se puede duplicar el efecto sobre la tensión de salida, produciéndose un incremento de la sensibilidad (figura 7.10a). La expresión de la tensión de salida resulta ser, de forma aproximada:

$$v_{AB} = -\frac{k\varepsilon}{2} E \tag{7.10}$$

Si se puede trabajar con estados tensionales "opuestos" pero iguales en valor absoluto es posible disponer cuatro galgas activas, dos trabajando a tracción y dos a compresión con lo que se duplica la sensibilidad del sistema (figura 7.10b) y la tensión de salida será:

$$v_{AB} = -k\varepsilon E \tag{7.11}$$

con cuatro veces más sensibilidad que en el montaje original y dos veces más sensibilidad que en la topología anterior. Claro está que esta posibilidad sólo se podrá llevar a cabo cuando existan estados tensionales opuestos en la zona de medida que lo permitan. Aun en ese caso, la utilización de las galgas debe hacerse con sumo cuidado para conseguir que realmente dos de ellas soporten una elongación positiva y las otras dos la soporten negativa y del mismo valor.

Sea cual sea la topología empleada, la salida del circuito en puente incluye un error de linealidad que se incrementa con la elongación de las galgas con lo que la precisión del sistema será mayor en la medida en que los esfuerzos —y por ello, las elongaciones— sean pequeños. En la figura 7.11 se representa el error de linealidad en función de la elongación para un puente con una sola galga activa.

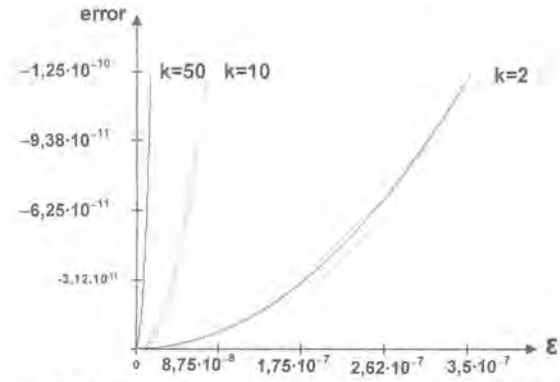


Figura 7.11. Error de linealidad en función de la elongación.

EJERCICIO 7.1

Determine la tensión de salida de un puente con una única galga activa de 160,000 Ω y resistencias de ese mismo valor alimentado a 5,00000 V cuando está sometido a una deformación de 500 με. Considérese que el factor de galga es K = 2,00000. ¿Qué error se cometería si se considerase el funcionamiento del puente como totalmente lineal? ¿Qué pasaría si se tuviese una deformación 10 veces mayor?

SOLUCIÓN:

Utilizando la expresión completa (7.9) tenemos que:

$$v_{AB} = \frac{2 \cdot 500 \cdot 10^{-6}}{4} \frac{1}{1 + \frac{2 \cdot 500 \cdot 10^{-6}}{2}} 5 = 1,24937 \cdot 10^{-3} \text{ V}$$

Si usamos la expresión "aproximada" que aparece en (7.9), el valor de la tensión de salida del puente sería de $1{,}25000 \cdot 10^{-3}$ V, es decir, un error de sólo 0,0500%. Con una elongación diez veces mayor, la tensión de salida real sería de:

$$v_{AB} = \frac{\frac{2 \cdot 5000 \cdot 10^{-6}}{4}}{1 + \frac{2 \cdot 5000 \cdot 10^{-6}}{2}} 5 = 0{,}0124378 \text{ V}$$

mientras que la expresión aproximada proporcionará un valor de 0,0125000 V, provocando un error de 0,497% , casi 10 veces más que antes.

La tensión de salida del puente suele ser muy pequeña y será necesario una etapa de amplificación posterior para obtener unos niveles razonables. Ese amplificador se construirá mediante alguna topología diferencial ya que ese es precisamente el tipo de salida que proporciona el puente de medida. Con los valores bajos que se están manejando, la selección del amplificador debe tener muy en cuenta los parámetros de continua tales como las tensiones y corrientes de desviación y sus correspondientes derivas. Otro factor a tener muy en cuenta será la CMRR ya que el sistema va a estar trabajando con un elevado modo común en relación con la tensión de salida del puente. Como comprobación de este término, en el ejercicio 7.1 la relación entre la tensión de salida del puente en el caso de máxima elongación y su modo común es de –46 dB.

Además de la importancia implícita del modo común que es capaz de producir un error en la salida del puente una vez amplificada, puede presentar el peligro potencial de saturación de las etapas intermedias de los amplificadores de instrumentación tal y como se estudió en el capítulo 3. En la figura 7.12 se muestra un puente de galgas con el correspondiente amplificador de instrumentación.

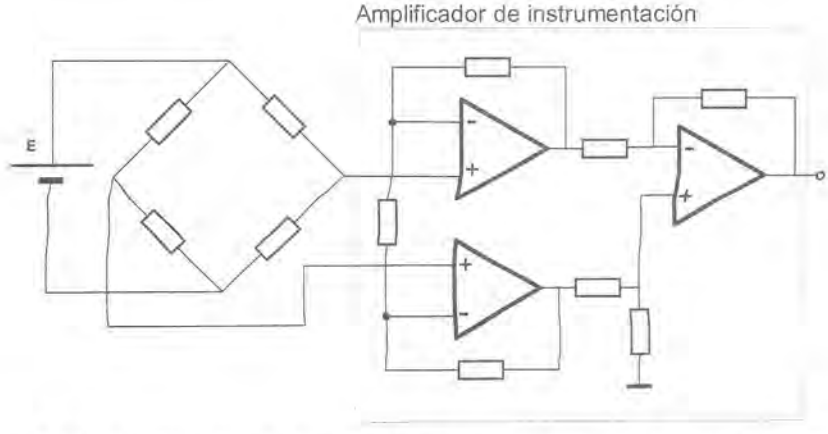


Figura 7.12. Amplificación de la señal de salida de un puente de galgas empleando un amplificador de instrumentación de tres operacionales.

EJERCICIO 7.2

Se pretende construir un sistema de medida de deformaciones con un puente formado por galgas extensométricas de 160 Ω (una de ellas activa con K = 2,000) alimentado a 5 V y que sea capaz de medir deformaciones desde 0 hasta 5000 με y proporcionar una salida de tensión desde 0 hasta 5 V, respectivamente. Se usará un amplificador de instrumentación INA 118 (ver hoja de datos en el CD). Teniendo en cuenta la CMRR del amplificador, ¿qué error se produciría en la salida? ¿Cómo se podría eliminar?

SOLUCIÓN:

Supongamos, en principio un esquema como el de la figura 7.12 y asumamos el puente como perfectamente lineal y equilibrado. En estas condiciones, a la salida del puente tendremos un margen de tensiones comprendidas entre 0,000 V y 12,5 mV. La ganancia necesaria para llevar la salida a 0-5 V será de $5/12{,}5 \cdot 10^{-3} = 400$ lo que es perfectamente alcanzable con un amplificador de instrumentación según se ha visto en el capítulo 3. Acudiendo a las hojas de características, el fabricante nos dice que la configuración del amplificador es la mostrada en la figura 7.13 y que su ganancia se consigue con una resistencia R_G cuyo valor viene determinado por la expresión:

$$A = 1 + \frac{50K}{R_G}$$

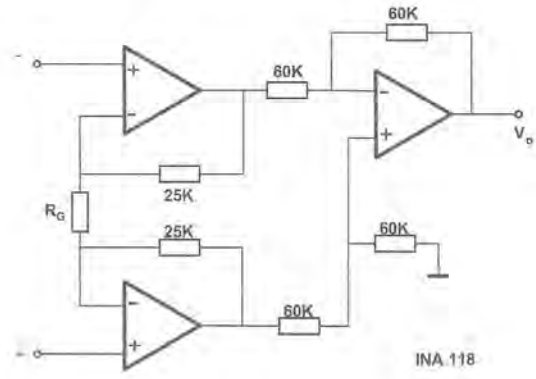


Figura 7.13. Amplificador de instrumentación de tres operacionales INA 118.

Para nuestro caso, R_G resulta ser de 125,3 Ω que habrá que aproximar al valor estándar más cercano o incluir un potenciómetro si se desea un ajuste exacto de ganancia con los problemas asociados a los ajustes en un circuito real que ya se han comentado. En la serie E96 de la EIA (véase Anexo en el CD) el valor más próximo es de 126 Ω y el 1% de tolerancia, luego este será el valor utilizado. La ganancia real del amplificador estará comprendida entre el valor que corresponda para resistencias de 126 ± 1%, es decir, $394 < A < 402$.

Una comprobación necesaria nos lleva a que las salidas de los amplificadores de entrada no se saturan para los valores de tensiones previstos en la entrada con lo que es de esperar un correcto funcionamiento del sistema.

Según los datos del fabricante, la CMRR en continua de este amplificador es generalmente un poco mayor de 120 dB para una ganancia entre 100 y 1000 como es nuestro caso, con un valor mínimo garantizado de 107 dB para A = 100. Supongamos este como valor de diseño (estamos en un caso más desfavorable del real) y como quiera que en la entrada tenemos un modo común de 2,5 V y una señal a fondo de escala de 12,5 mV, a la salida del amplificador habrá 5 V de tensión debidos a la señal y 4,5 mV debidos al modo común que suponen un error inferior al 0,1%.

El esquema final del diseño incluye una referencia estable de tensión como la LM136 (véanse características en el CD adjunto) que proporciona los 5 V necesarios para alimentar el puente a partir de una tensión de 10 V menos estable que alimenta el amplificador seleccionado[3] tal y como se muestra en la figura 7.14.

[3] Algunos amplificadores de instrumentación están previstos para trabajar con un puente de medida a la entrada e incorporan una referencia estable que proporciona la alimentación al puente incrementando el nivel de integración del diseño (consultar datos de fabricante en el CD adjunto).

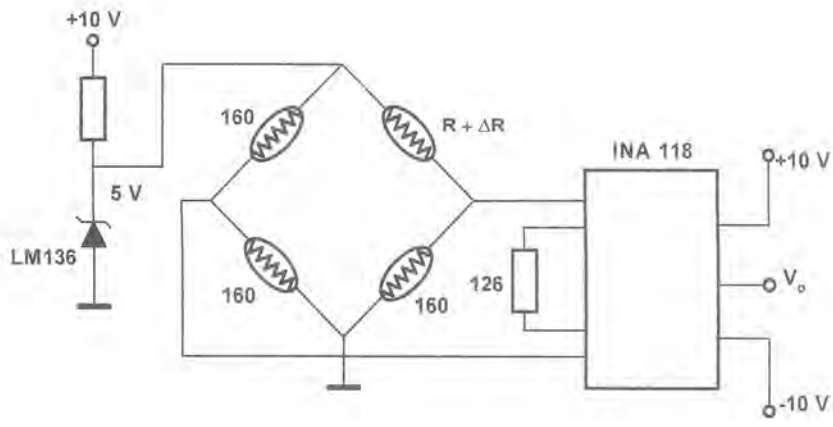


Figura 7.14. Uso del INA 118 como amplificador para un puente de galgas.

7.5 Utilización de los circuitos de medida

La utilización de los puentes de medida con galgas no es tan sencilla como pueda parecer hasta ahora puesto que habrá que tener en cuenta diversas circunstancias que permitan hacer que la medida sea correcta. Como aspectos más importantes cabe citar:

a) El **equilibrio** del puente que permite fijar su valor a cero en un punto concreto de medida.

b) La **calibración** del puente que permite obtener otro punto de la curva de medida y que permitirá fijar la ganancia del sistema para que se ajuste a las necesidades.

c) El **cableado** del sistema para reducir efectos no deseados y llevar a cabo una medida correcta, ajena (en la medida de lo posible) a interferencias externas.

Equilibrio del puente

Cuando se instala el puente y la galga (o galgas) activa no está sometida a esfuerzo alguno, el puente debería estar perfectamente equilibrado y proporcionar tensión nula en su salida. Esto no será así en la totalidad de los casos porque las resistencias no tienen los valores exactos o porque su comportamiento térmico no sea exactamente igual. Sea cual sea la causa, aparecerá una pequeña tensión en la salida que puede corregirse mediante la inclusión de un ajuste externo como el que se muestra en la figura 7.15.

Actuando sobre el potenciómetro se puede introducir una tensión adicional en el punto A del puente y conseguir que su salida sea perfectamente nula. Las resistencias empleadas y el potenciómetro deberán ser de un valor suficientemente alto como para no afectar seriamente al trabajo del puente lo que no suele ser un grave problema habida cuenta de lo bajo de los valores de las resistencias del puente.

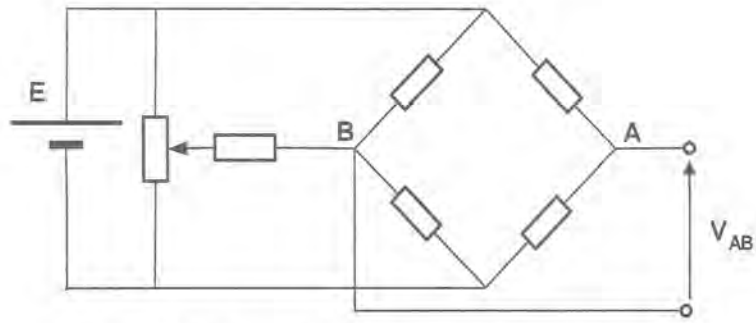


Figura 7.15. Circuito para el equilibrado de un puente.

Calibración

Normalmente los sistemas como el anterior (figura 7.15) suelen precisar calibración, es decir, situar un valor de deformación conocido para evaluar la salida del conjunto y comprobar si se acerca o no al valor previsto. Como quiera que situar un estado de deformación determinado puede no ser algo sencillo de hacer, se suele recurrir a procedimientos de simulación de deformación mediante la adición de una resistencia que desequilibre el puente como si se tratase de un esfuerzo aplicado sobre la galga. En la figura 7.16 se muestra una de estas técnicas conocida como *calibración mediante shunt*.

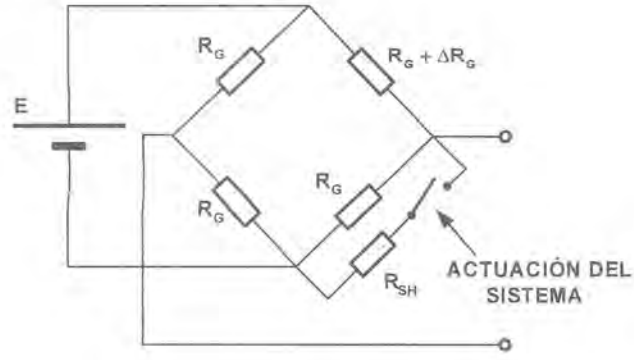


Figura 7.16. Calibración mediante *shunt* para simular una deformación determinada.

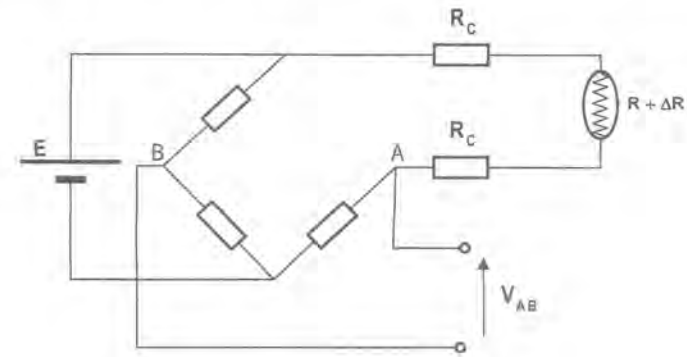
La idea consiste en la inclusión de una resistencia en paralelo R_{SH} con la resistencia del puente que esté en la misma rama que la galga activa. La deformación que simula esa resistencia viene dada por:

$$\varepsilon = \frac{R_G}{K\left(R_G + R_{SH}\right)} \tag{7.12}$$

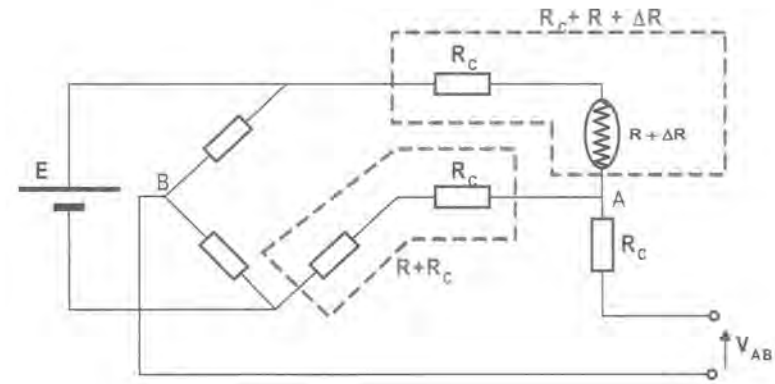
La resistencia que se va a utilizar es de muy bajo valor con lo que habrá que ser extremadamente cuidadoso al elegir el interruptor de cierre ya que sus contactos deberán introducir una resistencia adicional muy baja en cualquier momento de su vida útil. Para estos casos se recomienda el uso de relés encapsulados al vacío con contactos bañados en oro que previenen su degradación por oxidación y presentan un valor óhmico muy bajo.

Cableado del puente

Al utilizar una galga extensométrica para medida y conectarla al puente nos encontraremos en la mayoría de los casos con que el puente no puede estar en el mismo lugar que la galga (o galgas) y, en un buen número de casos, la distancia entre ambos puede ser grande como para que las resistencias propias del cableado afecten a la medida (figura 7.17a). Claro está que este efecto puede compensarse procediendo al equilibrado del puente antes de utilizarlo para medir, pero lo que no puede evitarse es el efecto de los cambios en esas resistencias debidos a la temperatura que tendrán el mismo efecto que los cambios resistivos en la galga producidos por la deformación que se pretende medir.



(a)



(b)

Figura 7.17. Efecto del cableado sobre el puente de medida: (a) los cables introducen una resistencia parásita; (b) la medida a tres hilos minimiza el efecto del cableado.

Para compensar los efectos de los cambios en las resistencias del cableado R_C se puede proceder a efectuar una medida a tres hilos (figura 7.17b) que permite que los cambios en R_C se produzcan tanto en la parte inferior del puente como en la superior y no tengan influencia en el valor de la tensión del punto A. Claro está que el tercer hilo también sufriría cambios de resistencia debidos a cualquier factor pero esto no tendrá mayor importancia puesto que el amplificador que leerá la tensión del puente tendrá una impedancia muy elevada y la influencia del cableado será nula.

7.6 Aplicaciones

Las aplicaciones de las galgas extensométricas pueden clasificarse en dos grandes grupos: en primer lugar, las que suponen la medida directa de un estado de deformación en una superficie que permitirá conocer su estado tensional con las restricciones ya establecidas; en segundo lugar, aquellas en las que el estado de deformación es una variable intermedia consecuencia de la variable que se pretende medir. A continuación, se tratarán estas dos aplicaciones.

7.6.1 MEDIDA DE ESTADOS DE DEFORMACIÓN

Medidas estáticas

Se trata de medidas en las que se supone que los cambios de los estados de deformación ocurren con lentitud y, por ello, no será preciso cuidar la respuesta en frecuencia de los amplificadores, pudiéndose utilizar filtros paso bajo con frecuencias de corte bajas para limitar la presencia de señales de ruido de alta frecuencia que podrían aparecer. Este tipo de medidas corresponden a sistemas de monitorización de elementos de sustentación en construcción o en cualquier tipo de máquina que no esté sometida a estados vibracionales o de fatiga.

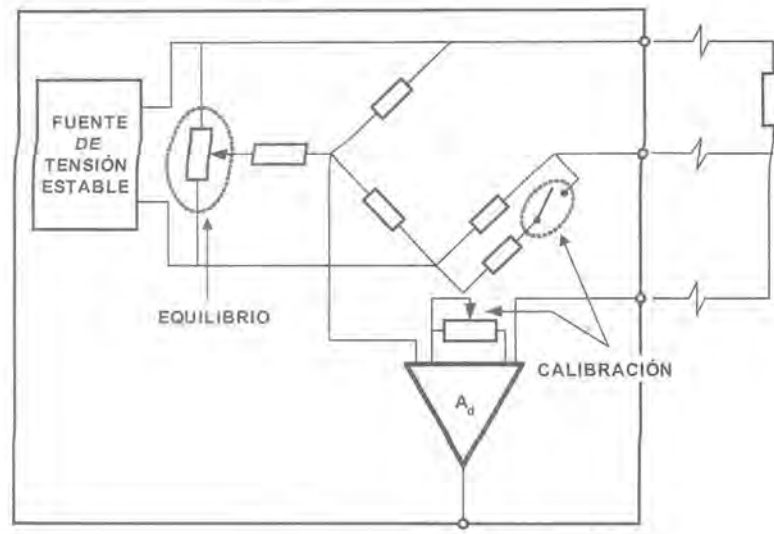


Figura 7.18. Circuito de medida con una galga activa que incluye el equilibrado del puente, el ajuste de la ganancia y de un punto de deformación.

Dentro de los sistemas de medida de estados de deformación mediante galgas extensométricas se tienen los casos de sistemas en que se pretende medir un único estado con una variable y que incluirán un único puente con los accesorios adicionales para el ajuste del cero y la calibración. Este tipo de sistema corresponde directamente a lo que se ha descrito hasta este momento y supone una amplificación posterior hasta llevar la tensión de salida a los valores que se desee. En la figura 7.18 se puede observar la topología de este tipo de sistemas.

Sin embargo, la situación más normal es la medida de los estados de deformación en varios puntos simultáneamente lo que implica el uso de diferentes canales para la medida. En algunos casos, tales como los sistemas de monitorización de los procesos de construcción de grandes edificios o infraestructuras viarias, el número de variables de deformación monitorizadas puede llegar a ser muy grande (cientos de canales) y los equipos de medida pueden realizarse de dos formas:

a) Mediante un único amplificador y un multiplexador que vaya recorriendo los distintos puentes de medida de forma secuencial (figura 7.19a). Se tiene así un ahorro de equipamiento ya que los amplificadores aislados de instrumentación y las fuentes de alimentación ultra-estabilizadas suelen tener un coste alto. El inconveniente está en la imposibilidad de realizar medidas simultáneas aunque esto es menos importante en medidas estáticas.

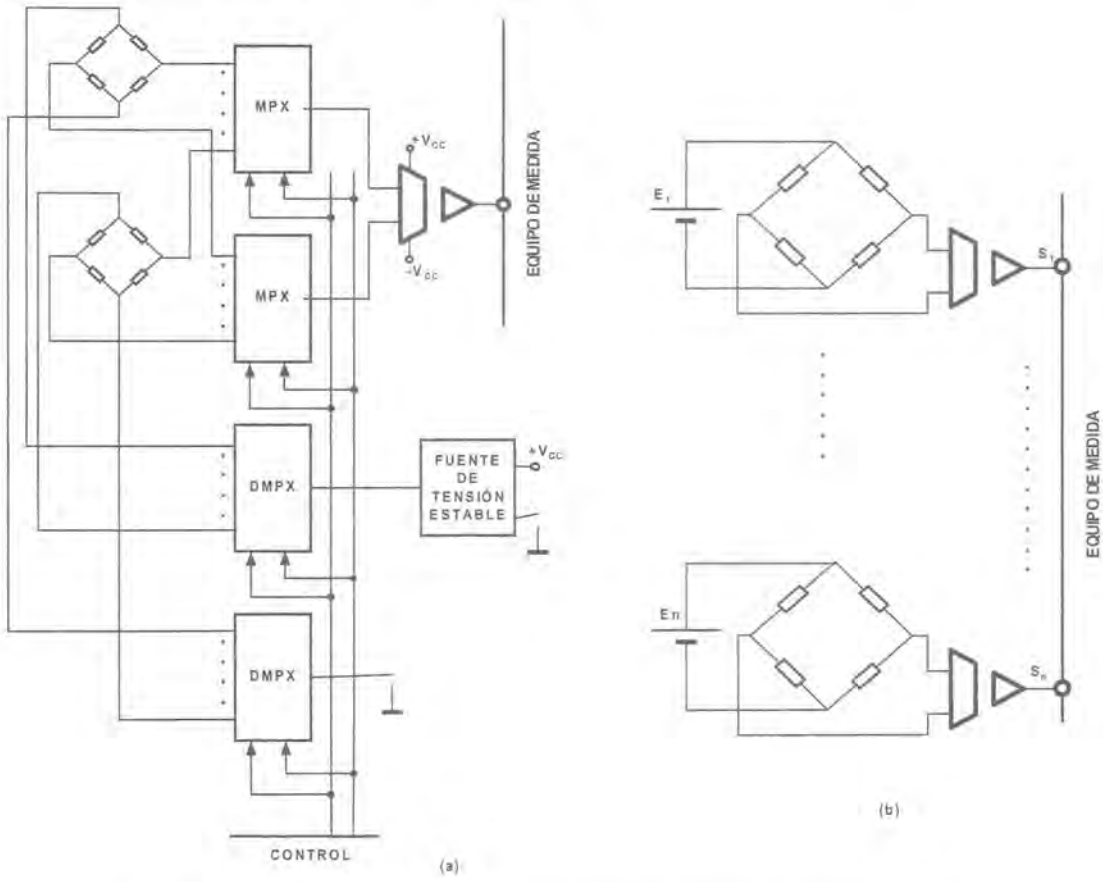


Figura 7.19. Circuitos de medida multicanal con galgas extensométricas: a) caso de un conjunto de puentes de galgas alimentado desde una única fuente de tensión; b) circuito con alimentaciones independientes para cada canal.

b) Mediante fuentes y amplificadores independientes para cada puente de medida con un coste muy alto, pero una garantía de simultaneidad total y la posibilidad de realizar ajustes particulares para cada uno de ellos (figura 7.19b).

La primera solución se suele emplear cuando son muchos los puentes a medir mientras que la segunda se usa —en general— para casos con menos variables. También es posible una solución intermedia entre estas

dos a costa de multiplexar grupos de puentes con un amplificador para cada grupo, lo que puede ser una solución de compromiso entre prestaciones y coste.

Medidas dinámicas

Cuando el objetivo es conocer la evolución en el tiempo (o en frecuencia) de los estados de deformación en uno o varios puntos de una pieza o sistema mecánico, es necesario realizar medidas dinámicas que recojan este tipo de información. En estas condiciones, será preciso la utilización de sistemas como los de la figura 7.19 aunque no suelen ser muchas las variables que se miden simultáneamente. Las galgas tienen (o pueden tener) buenas características dinámicas por lo que el cuello de botella de este tipo de sistemas suele estar en el producto ganancia-ancho de banda del amplificador que se convierte en un equipo de alto coste [7.5]. Con vistas a optimizar la relación S/N se incluirán filtros a la salida del amplificador para limitar el ancho de banda al estrictamente necesario para el trabajo a realizar y evitar el procesamiento de tensiones fuera de las frecuencias deseadas. En la figura 7.20 se muestra un ejemplo de este tipo de sistemas.

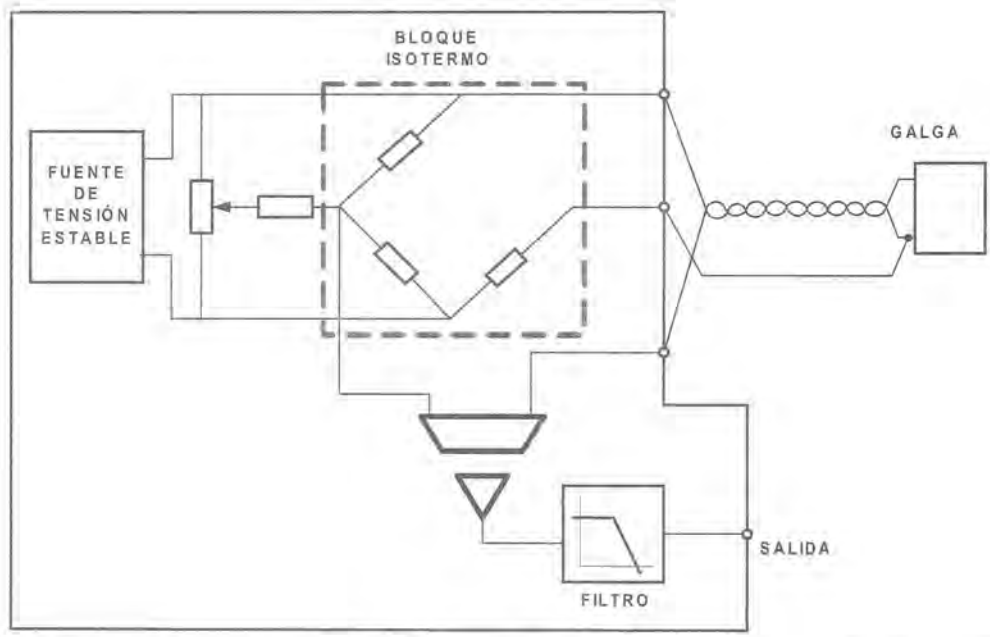


Figura 7.20. Sistema para medidas dinámicas con galgas extensométricas que incluye bloques isotermos para las galgas pasivas (para evitar que se produzcan diferencias térmicas entre las galgas pasivas con los consiguientes errores) y una etapa de filtrado para mejorar la relación S/N.

Los sistemas comerciales de este tipo suelen incluir filtros de frecuencia de corte variable, adaptables a las necesidades de cada caso, amplificadores aislados con las prestaciones para ajustar el cero y la calibración mediante resistencia *shunt* y fuentes ultra-estabilizadas para alimentar el puente.

A la hora de diseñar estos sistemas, se deberá tener en cuenta el tipo de aplicación a la que van destinados con lo que, si se pretende realizar medidas de vibración, prestaremos especial atención a los parámetros de alterna de los amplificadores utilizados, mientras que, en medidas de fatiga, de impacto o en la que se mezclen las condiciones estáticas y dinámicas, habrá que considerar todas las características de los amplificadores (véanse capítulos 2 y 3) con lo que el problema puede llegar a complicarse en gran medida.

7.6.2 MEDIDAS DE OTRAS VARIABLES CON GALGAS EXTENSOMÉTRICAS

Hay muchas variables que pueden producir estados de deformación en una pieza con lo que las galgas extensométricas se pueden usar para medir esta variable intermedia y, así, poder medir la original. Los casos pueden ser muy variados y se deja a la imaginación del lector una gran cantidad de aplicaciones que serían factibles, citando aquí dos de las aplicaciones más importantes: la medida de peso y la medida de presión.

Células de carga (*load cells*)

Las células de carga (o celdas de carga) son piezas mecánicas con configuración muy variada [7.4] que se han diseñado de tal forma que en algunas partes de su estructura mecánica, el peso aplicado se transforma en una deformación proporcional que se mide con dos o cuatro galgas mediante topologías como las de la figura 7.10. En la figura 7.21 se muestran algunas realizaciones de células de carga similares a las que están disponibles en el mercado y que pueden medir desde algunos kilogramos hasta decenas de toneladas.

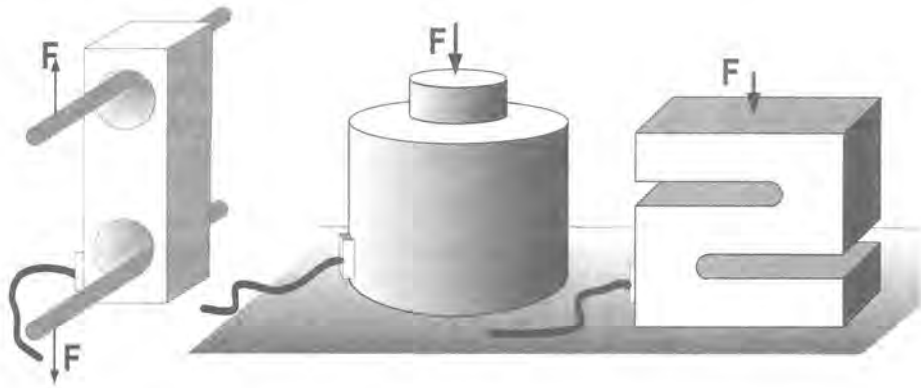


Figura 7.21. Celdas de carga para medida de fuerza que usan como dispositivo sensor una o varias galgas extensométricas.

Las medidas de peso suelen ser estáticas con lo que los sistemas de amplificación obedecen a las directrices expresadas en el apartado 7.6.1 sin especiales condiciones. La selección de las células de carga se hace en función del lugar en que se van a situar, el tipo de trabajo al que van a ser sometidas y las características estáticas y dinámicas típicas de cualquier sensor. En la tabla 7.4 se proporcionan los criterios y los tipos de células que se pueden encontrar en el mercado.

Criterio	Tipos de célula	Aplicaciones
Tipo de trabajo	Tracción	Medida de peso
	Compresión	Medidas on-line
	Tracción/compresión	Uso general
	Fatiga	Ensayos dinámicos
	Impacto	
Margen de medida	Microcélulas de carga	Alta precisión
	Margen amplio	Uso general
Comportamiento dinámico	Fatiga	Sistemas sometidos a fatiga
	Alta velocidad	Vibración Ensayos dinámicos

Tabla 7.4

Medida de presión

También es posible la medida de presión mediante un sistema de membrana que transforme la diferencia de presión entre sus dos caras en una deformación que se mide mediante un puente de cuatro galgas extensométricas en el que dos trabajan a tracción y dos a compresión, tal y como se muestra en la figura 7.22.

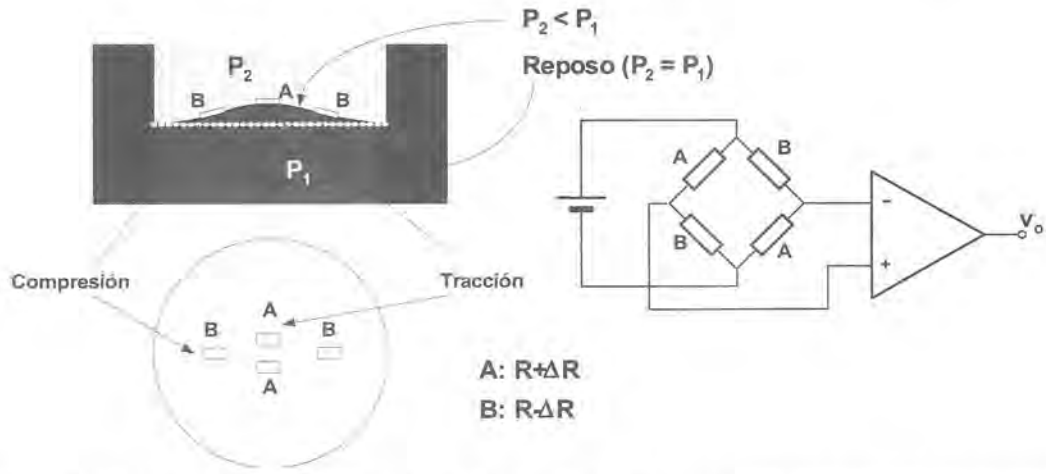


Figura 7.22. Medida de presión diferencial mediante un diafragma cuya deformación es recogida por un puente de cuatro galgas extensométricas activas.

La dificultad de este tipo de sistemas está en la previsible no-linealidad entre la diferencia de presiones y el estado de deformación que obliga a un diseño muy cuidadoso de la membrana, en cuanto a materiales (metálicos, cerámicos, plásticos, etc.) y morfología, y una selección correcta de los lugares en los que se van a situar las galgas de medida. De todas formas, la no linealidad se puede paliar en gran medida mediante un pequeño calculador incluido que incorpore una tabla de calibración del conjunto.

El sistema mostrado en la figura 7.22 permite la medida de presiones diferenciales pero también es posible medir presiones absolutas con sistemas similares en los que se ha fijado una atmósfera de referencia de presión conocida y temperatura controlada.

Resumen

En el presente capítulo se ha hecho un breve recorrido por el mundo de las galgas extensométricas, tratando de comentar el principio de funcionamiento basado en el cambio de la resistencia efectiva de un material cuando es sometido a una deformación así como los materiales más habituales para la fabricación de estos dispositivos. Aquí se han diferenciado claramente las galgas metálicas de las semiconductoras, las primeras con mayor precisión pero menor sensibilidad que las segundas aunque estas últimas están aún en una fase de desarrollo. También se ha comentado brevemente la presentación comercial de las galgas, sus configuraciones más típicas y las particularidades de uso y aplicación en la medida de esfuerzos en piezas mecánicas.

Como aspecto básico de todo dispositivo sensor tenemos el circuito medida: para las galgas se usa siempre un circuito en puente aunque se pueden incluir una, dos o —incluso— cuatro galgas activas con lo que se consigue aumentar la sensibilidad del conjunto. Para este circuito de medida se han analizado sus porme-

nores (en especial, en lo que se refiere a este caso concreto), exponiendo las técnicas de equilibrado y calibración, así como las precauciones a tener en cuenta para hacer que la medida sea correcta, tales como los aspectos térmicos, el cableado a tres hilos o la importancia de una buena alimentación. La salida de un puente debe amplificarse —máxime cuando las galgas de medida son metálicas— por lo que se usará un circuito amplificador de instrumentación de tres operacionales.

Los circuitos de medida con galgas suelen emplear un elevado número de sensores para hacer gran cantidad de medidas sobre una estructura o una pieza mecánica por lo que deben configurarse auténticos sistemas de instrumentación multicanal cuyas configuraciones básicas también se han presentado.

Para finalizar se han incluido algunas aplicaciones muy habituales en las que las galgas se usan como sensor final para hacer medidas diferentes a las que corresponden a la elongación: es el caso de determinados sensores tales como los de fuerza y peso o los de presión que —mediante el diseño de una pieza mecánica— traducen la variable correspondiente en una deformación que se mide con galgas como es el caso de las celdas de carga o los sensores de presión diferencial de diafragma.

Problemas propuestos

7.1 Un puente formado por cuatro galgas de 160,0 Ω y K = 2,000 con una de ellas activa trabajando a tracción está alimentado a 5,000 V y va a soportar una elongación máxima de 5000 με. Calcule el margen de tensión de salida.

7.2 El mismo puente del problema 7.1 posee dos galgas activas que trabajan con la misma elongación máxima. ¿Cuál es ahora su margen de tensiones en la salida? Compárela con el problema anterior.

7.3 Si desease amplificar la señal procedente de los puentes de los problemas 7.1 y 7.2, ¿cuáles serían las especificaciones del amplificador para conseguir que la tensión de salida del sistema se encuentre comprendida en el margen 0 a 10 V?

7.4 Reconsidere el problema anterior si el puente utilizase galgas de semiconductor con K = 100,0.

7.5 ¿Qué impedancia de salida presenta un puente de galgas metálicas 160,0 Ω en reposo?

7.6 Se tiene un puente con cuatro galgas de 160,0 Ω y K = 2,000. La medida se hace con una de las galgas que trabaja a compresión con una deformación máxima de 2000 με separada del puente una distancia de 50 cm. Suponiendo que el cable empleado para el cableado tiene una resistencia por unidad de longitud de 0,5 Ω/m, ¿qué error a fondo de escala se produce por el cableado?

7.7 En el problema anterior considérese un circuito de equilibrado del puente que elimina el error del cableado. ¿Qué ocurre si la temperatura modifica la resistencia de los cables de conexión en un 5%?

7.8 Se dispone de un puente con galgas de 350,0 Ω de tipo metálico con K = 2,000 de las cuales una es activa y trabaja a compresión. Determine el error de linealidad obtenido si se trabaja con elongaciones máximas de 2000 με.

7.9 Se dispone de un puente con galgas de 350,0 Ω de semiconductor con K = 50,00 de las cuales una es activa y trabaja a compresión. Determine el error de linealidad obtenido si se trabaja con elongaciones máximas de 2000 με.

7.10 Compare los resultados de los problemas 7.8 y 7.9 en relación con la gráfica de la figura 7.11.

7.11 Se tiene un puente con galgas metálicas de 1000 Ω y K = 2,000 (sólo una es activa) sometido a una deformación máxima de 3000 με. ¿Cuál sería la deformación que se debería producir sobre un puente de galgas de igual valor óhmico pero K = 50,00 para que la salida sea la misma que en el caso anterior a fondo de escala?

7.12 Compare los dos puentes respecto a sensibilidad y error de linealidad. ¿Cuál sería el margen de uso de cada uno de ellos?

7.13 Utilizando un amplificador INA 118 se amplifica la señal del puente del problema 7.2. Considerando que su CMRR es de 120 dB, determine el error debido al modo común.

7.14 Reconsidere la cuestión anterior si el puente se alimenta a 10,00 V.

8

Termistores y fotorresistencias

8.1 Introducción
8.2 Termistores: NTCs
8.3 Termistores de coeficiente de temperatura positivo (PTCs)
8.4 Fotorresistencias (LDRs)

8.1 Introducción

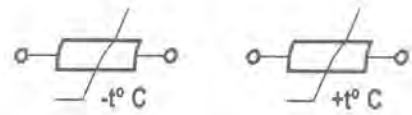
Los **termistores** son sensores de temperatura de tipo resistivo. El nombre de termistor nace de la contracción de las palabras inglesas "*thermal*" y "*resistor*" (resistencia sensible a la temperatura). Los termistores se dividen en dos grupos atendiendo al signo del coeficiente de temperatura de la resistencia (α): **NTC** (*Negative Temperature Coefficient*) que presentan un coeficiente de temperatura negativo y **PTC** (*Positive Temperature Coefficient)* con un coeficiente de temperatura positivo.

El capítulo concluye con el estudio de las fotorresistencias que son sensores resistivos para la medida y detección de radiación electromagnética.

8.2 Termistores: NTCs

Principio físico de funcionamiento

Las NTCs son resistencias de material semiconductor cuya resistencia disminuye cuando aumenta la temperatura, y están constituidas por una mezcla de óxidos metálicos. Generalmente se utilizan combinaciones de Ni-Mn-O, Ni-Cu-Mn-O y Ti-Fe-O. El mecanismo de conducción en este tipo de materiales semiconductores es muy complejo, al menos comparado con lo que ocurre en un simple cristal de silicio. Básicamente, el incremento de temperatura aporta la energía necesaria para que se incremente el número de portadores capaces de moverse, lo que lleva a un incremento en la conductividad del material. Los símbolos eléctricos de los termistores se indican en la figura 8.1.



(a) (b)

Figura 8.1. (a) Símbolo eléctrico de una NTC, el signo menos indica que la sensibilidad es negativa, la línea que corta al dispositivo alude a que la resistencia cambia con la temperatura; (b) símbolo eléctrico de una PTC.

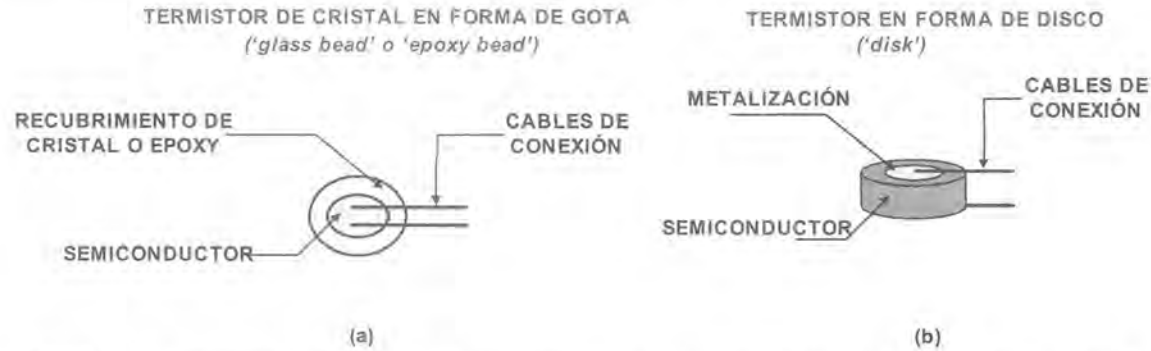
Tecnología de fabricación

Para fabricar una NTC se parte de la mezcla de los óxidos metálicos hasta formar un polvo homogéneo. El polvo se comprime y se sinteriza a alta temperatura dándole la forma deseada. A esta estructura se le conectan los terminales y se sella mediante una capa aislante de "*epoxy*" o cristal que le proporciona protección y la aísla eléctricamente. El tipo de óxidos seleccionados y su proporción, junto a la atmósfera y temperatura de sinterizado se controla cuidadosamente ya que determinará la resistividad del material y el coeficiente de temperatura del termistor.

Las NTC se clasifican en dos grupos atendiendo a la forma en que se unen los cables al cuerpo de la NTC:

- Termistores tipo ***bead*** o **gota**
- Termistores **metalizados**

En los **termistores** de tipo ***bead*** o **gota**, los cables se sinterizan junto al cuerpo del termistor (figura 8.2a). En los **termistores metalizados** los cables se conectan a la superficie del termistor que previamente se ha metalizado (figura 8.2b). Los termistores metalizados pueden adoptar diversos tamaños y formas (puede ver en el CD el aspecto de diversos tipos de termistores).



(a) (b)

Figura 8.2. Ejemplos de termistores: (a) termistor de tipo *bead*, los cables de conexión forman parte del cuerpo del termistor; (b) termistor metalizado en forma de disco, los cables se conectan a las superficies metalizadas del termistor.

Los termistores de tipo *bead* son más estables y soportan mayores temperaturas (hasta 550°C) que los metalizados, por el contrario, su tolerancia y su coste son mayores. Las tolerancias en la fabricación de termistores convencionales suelen ser elevadas (±20% en los *bead* y de ±5% en los metalizados). Algunos termistores especiales presentan tolerancias más reducidas; por ejemplo, en algunos se ajusta la resistencia del termistor en un punto concreto de temperatura (termistores *point-matched*).

Los termistores no son tan estables como las RTDs (1%/año para los de baja calidad); para mejorarla suele controlarse la calidad de los materiales utilizados en su fabricación y someterse a las unidades a un proceso de envejecimiento a elevadas temperaturas.

8.2.1 CARACTERÍSTICA R-T DE UNA NTC

La característica R-T de una NTC se muestra en la figura 8.3. Se observa que la relación entre la resistencia y la temperatura no es lineal, sobre todo, cuando se considera un margen de temperatura amplio. Por otro lado, la sensibilidad (pendiente) es muy grande a bajas temperatura y va disminuyendo conforme aumenta ésta. Una **sensibilidad** alta es una característica muy deseable en cualquier sensor; de hecho, es la mayor ventaja de los termistores frente a otros sensores de temperatura. En una NTC típica, el coeficiente de temperatura de la resistencia (α) es unas diez veces mayor que en los metales, del orden de -3 a $-6\%/^{\circ}C$. Para contrastar este comportamiento se dibuja en la figura 8.3 la curva R-T de una RTD.

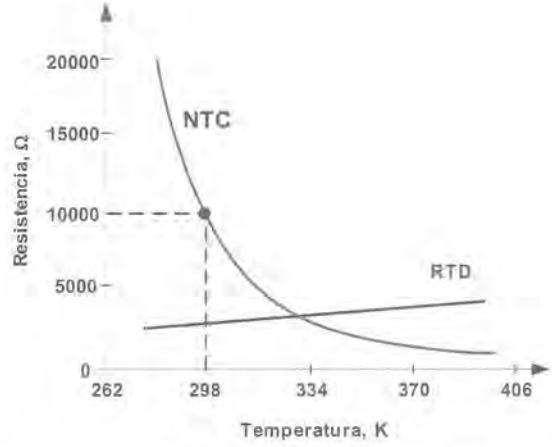


Figura 8.3. Relación entre resistencia y temperatura de una NTC de 10 kΩ a 25°C (298 K) y de una RTD.

En la figura 8.3 se indica también la resistencia nominal de una NTC que hace referencia a su valor resistivo a una temperatura de referencia, generalmente 25°C (298 K). Los valores de resistencia nominal más comunes varían entre 10 Ω y 20 MΩ. Las NTCs miden temperaturas absolutas, con un valor máximo de 500°C (aunque existen nuevos materiales que permiten llegar a 1000°C) y la mínima alrededor de −70°C. Su velocidad de respuesta no es muy elevada aunque pueden fabricarse termistores de tamaño muy reducido que presentan una baja inercia térmica, además, y a pesar de su tamaño, pueden conseguirse unidades con un alto valor resistivo que hace despreciable la influencia de los cables de conexión y facilita el acondicionamiento.

Las NTCs son componentes muy robustos, fiables, sensibles y económicos. Sus mayores inconvenientes son: su lentitud de respuesta, la presencia de grandes tolerancias en su fabricación, una relativa estabilidad y un campo de medida también algo limitado.

8.2.2 APROXIMACIONES DE LA CARACTERÍSTICA RESISTENCIA-TEMPERATURA

Modelo exponencial

La ecuación más sencilla que reproduce el comportamiento de un termistor ideal es:

$$R_T = R_0 e^{B\left(\frac{1}{T}-\frac{1}{T_0}\right)} \tag{8.1}$$

donde, R_T es resistencia del termistor a la temperatura T en Kelvin, B es el índice de sensibilidad del termistor en Kelvin y R_0 es la resistencia del termistor a la temperatura T_0.

Para evaluar los parámetros R_0 y B se necesitan dos puntos de calibración. Puede medirse la resistencia, por ejemplo, en $T_0 = 298$ K (25°C) y en $T_1 = 358$ K (85°C); si llamamos R_0 y R_1 a los valores resistivos obtenidos a esas temperaturas resulta:

$$B = \frac{1}{\frac{1}{T_0}-\frac{1}{T_1}} \ln \frac{R_0}{R_1} \tag{8.2}$$

Los fabricantes utilizan el parámetro B como una constante asociada al material con que se fabrica el termistor y suele variar entre 2000 a 6000 K.

De la ecuación (8.1) se puede obtener la temperatura:

$$T = \frac{BT_0}{B + T_0 \ln \frac{R_T}{R_{T_0}}} \tag{8.3}$$

y el coeficiente térmico:

$$\alpha = \frac{1}{R_T}\frac{dR_T}{dT} = -\frac{B}{T^2} \tag{8.4}$$

que depende de la temperatura.

La ecuación (8.3) puede utilizarse en un programa de un microcontrolador para expresar la temperatura conociendo el valor resistivo de la NTC. Este modelo, sin embargo, sólo es válido para un margen de temperaturas reducido. Para un termistor típico trabajando entre 0 y 50°C el error cometido está en torno a ±0,3°C.

8.2.3 LA NTC COMO ELEMENTO DE CIRCUITO

Cuando conectamos una NTC a un circuito eléctrico se le suministra energía que por efecto Joule se transforma en energía térmica (figura 8.4a). Parte de la energía suministrada es disipada al exterior y otra parte es absorbida, incrementándose su temperatura sobre la temperatura ambiente.

La ecuación de transferencia de calor que gobierna el proceso se modela por la siguiente ecuación:

$$P = V\,I = \delta(T - T_a) + c_p m \frac{dT}{dt} \tag{8.5}$$

donde T es la temperatura del termistor, T_a es la temperatura del medio que rodea al termistor o temperatura ambiente, δ es el coeficiente de disipación o conductancia térmica, m es la masa del termistor, $c_p m$ es la capacidad calorífica, V e I son la tensión y la corriente por el termistor y P es potencia eléctrica suministrada.

La solución de la ecuación diferencial anterior considerando P constante es:

$$T = T_a + \frac{P}{\delta}\left[1 - e^{-\frac{\delta}{c_p m}t}\right] = T_a + \frac{P}{\delta}\left[1 - e^{-\frac{t}{\tau}}\right] \tag{8.6}$$

De (8.6) se desprende que la temperatura del termistor se incrementa exponencialmente sobre la temperatura ambiente (figura 8.4b). Este incremento de temperatura se debe al **autocalentamiento** y es consecuencia de la energía absorbida. Se denomina **constante de tiempo térmica** al cociente entre su capacidad calorífica y su constante de disipación térmica, es decir, $\tau = c_p m/\delta$.

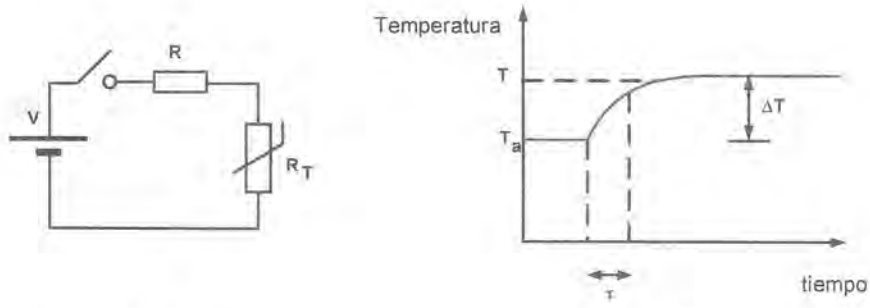


Figura 8.4. (a) La NTC como elemento de un circuito; (b) al cerrar el interruptor la temperatura en la NTC asciende pasando de T_a a T (τ es la constante de tiempo).

EJERCICIO 8.1

Se dispone de la siguiente información en un termistor: la capacidad calorífica es 1000 mJ/K y el coeficiente de disipación, 8 mW/K en aire. Calcule la constante de tiempo térmica en aire.

SOLUCIÓN:

$$\tau = c_p m/\delta = 1000 \text{ mJ/K} / 8 \text{ mW/K} = 125 \text{ s}$$

Pasado un tiempo se alcanza el equilibrio térmico, momento en el que la temperatura de la NTC se hace constante (dT/dt = 0). En equilibrio térmico, toda la potencia suministrada se cede al ambiente resultando:

$$P = V_T I_T = \delta(T - T_a) \tag{8.7}$$

donde V_T e I_T son la tensión y la corriente en régimen permanente.

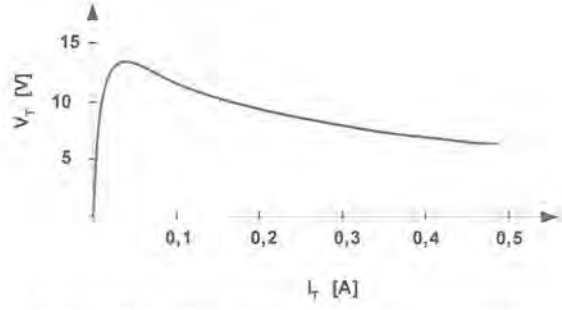


Figura 8.5. Característica V-I de un termistor.

Esta ecuación permite obtener la ***característica tensión-corriente*** *del termistor en condiciones estables* de temperatura (figura 8.5).

En el tramo ascendente de la curva V-I del termistor la potencia aplicada se considera despreciable (autocalentamiento despreciable) y es la zona de la característica que se utiliza en las aplicaciones termométricas. Si aumenta la corriente, llega a alcanzarse el máximo de tensión y, a partir de ese punto, la curva presenta una resistencia dinámica negativa. Los circuitos que trabajan en esta zona dan lugar a otro tipo de aplicaciones basadas en el autocalentamiento.

8.2.4 LA NTC COMO SENSOR DE TEMPERATURA

La medida de temperatura es la aplicación más común de las NTCs. Generalmente se selecciona una NTC de resistencia nominal alta y con un buen acoplamiento térmico con el objeto cuya temperatura se desea medir. Por otro lado, el autocalentamiento debe hacerse despreciable ya que introduce errores en la medida. El problema que debemos resolver es cómo obtener una magnitud eléctrica, por ejemplo, una tensión, que dependa de la temperatura de la forma más lineal posible, a pesar de que la relación R-T del termistor no sea lineal. En general, las técnicas de linealización pueden realizarse en dos dominios: el digital y el analógico. En el dominio digital, la estimación de la temperatura se realiza con independencia de la linealidad. Para ello se usa un computador que, utilizando el modelo del termistor, calcula el valor correspondiente de temperatura a partir de la medida del valor resistivo de la NTC. En el dominio analógico suele añadirse una resistencia en serie o en paralelo con el termistor para conseguir una respuesta que se aproxime más a la lineal. Veamos cómo se realiza el proceso de linealización:

a) Linealización mediante un divisor resistivo

Considérese el circuito de la figura 8.6a cuya tensión de salida se obtiene en la resistencia fija R. Esta disposición permite que la tensión de salida se incremente conforme lo hace la temperatura, además, puede incluirse en R la carga relacionada con el circuito externo de medida (cables e impedancia de entrada) consiguiendo resultados más precisos.

La tensión de salida es:

$$v_o(T) = v_i \frac{R}{R + R_T} \tag{8.8}$$

y la sensibilidad:

$$S(T) = \frac{dv_o(T)}{dT} = v_i \frac{RR_T}{[R + R_T]^2} \frac{B}{T^2} \tag{8.9}$$

En la figura 8.6b se muestra la salida del divisor normalizada respecto a la tensión de alimentación v_i. Se aprecia que la respuesta tiene la forma de una "s" (con un punto de inflexión) y es bastante lineal en un margen de temperatura limitado (±50 K).

El objetivo es encontrar un valor de R que proporcione una linealidad óptima en el margen de temperatura de interés y para una NTC dada. Un método analítico para determinar el valor de R consiste en hacer coincidir el punto de inflexión de la curva de salida (figura 8.6) con el punto medio de nuestro margen de medida, T_c (la tangente por el punto de inflexión coincide con la zona más lineal ya que la recta atraviesa la curva). La condición de punto de inflexión implica que debe anularse la derivada segunda de (8.8) en el punto medio del margen de medida. Resolviendo la ecuación para R resulta:

$$R = \frac{B - 2T_c}{B + 2T_c} R_{T_c} \tag{8.10}$$

donde R_{TC} es la resistencia del termistor en el punto central del margen de medida.

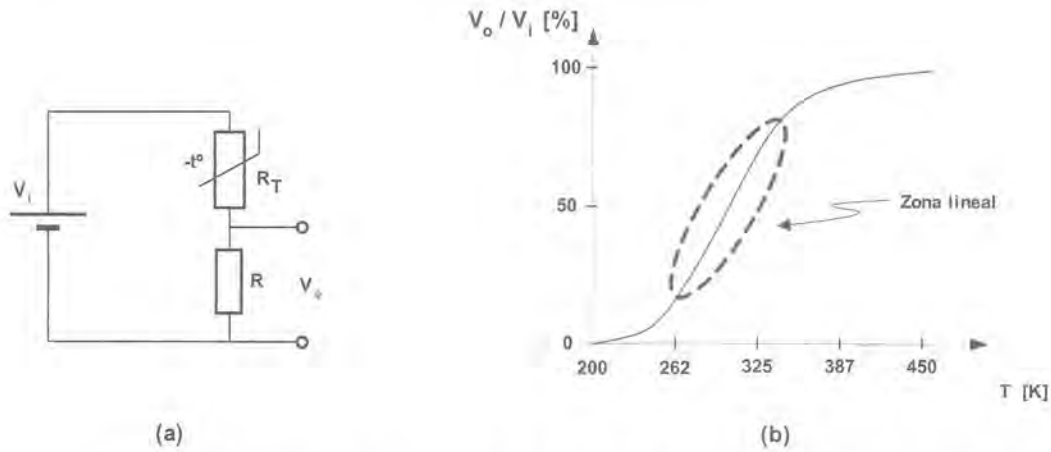


Figura 8.6. (a) circuito de linealización de una NTC basado en un divisor resistivo; (b) relación tensión-temperatura del divisor de tensión.

En relación con la tensión de alimentación del divisor, v_i, debe llegarse a un compromiso entre precisión, que lleva a elegir valores bajos de tensión para evitar el autocalentamiento y sensibilidad, que aconsejaría valores grandes. En general, se considera que el incremento de temperatura debido al autocalentamiento debe ser más pequeño que la precisión deseada en la medida. Teniendo en cuenta que el incremento máximo de temperatura por autocalentamiento (máxima potencia aplicada, $P_{máx}$) se producirá cuando la resistencia del termistor R_T sea igual a la resistencia fija del divisor, R, calculada anteriormente, se tiene:

$$P_{máx} = \frac{\left(v_{imáx}/2\right)^2}{R} \tag{8.11}$$

y por tanto:

$$v_{imáx} = \sqrt{4P_{max}R} = 2\sqrt{\Delta T \delta R} \qquad (P_{máx} = \delta \Delta T) \tag{8.12}$$

donde ΔT es el incremento de temperatura producido.

Una vez que se determina el valor de R y v_i, la sensibilidad en el punto medio del intervalo de medida se puede calcular sustituyendo (8.10) en (8.9):

$$S(T_C) = \frac{v_i}{B}\left(\frac{B^2}{4T_c^{\,2}} - 1\right) \tag{8.13}$$

y la recta de calibración que resulta es:

$$v_o(T) = v_o(T_c) + S(T_C)(T - T_C) \tag{8.14}$$

EJERCICIO 8.2

Realice un termómetro clínico para la medida de temperatura en el margen [35°C÷45°C] con un error máximo permitido de 0,1°C (1% del margen de medida). La salida deberá llevarse a un convertidor A/D que

admite tensiones de entrada entre 0 y 5 V. Utilice una NTC con las siguientes características: B = 4190 K, R = 100K @ 25°C, δ = 10 mW/°C.

SOLUCIÓN:

La temperatura central es (35°C + 45°C)/2 = 40°C , luego $T_C = 273 + 40 = 313$ K.

La resistencia del termistor, R_{TC} @ 40°C de acuerdo a la ecuación (8.1) es:

$$R_{TC} = 50980\ \Omega \text{ (escogemos } 51.000\ \Omega)$$

La resistencia del divisor según (8.10) es:

$$R = \frac{B - 2T_C}{B + 2T_C} R_{TC}$$

luego R = 37,7 kΩ.

Para el cálculo de la tensión de alimentación fijamos el error máximo de autocalentamiento a 0,01°C (diez veces menor que la precisión requerida). De la ecuación (8.12) resulta:

$$\Delta T = 0{,}01°C\ ;\ \delta = 10\ mW/°C = 0{,}01\ W/°C;\ R = 37{,}7\ k\Omega$$

luego: $V_{imáx} = 2\sqrt{\Delta T \delta R} = 3{,}88$ V. Elegimos $v_i = 2{,}5$ V por comodidad para obtenerla.

La tensión de salida real es:

$$v_o(T) = v_i \frac{R}{R + R_o e^{B\left(\frac{1}{T} - \frac{1}{298}\right)}}$$

Vamos a obtener la recta que linealiza la respuesta en el punto medio del margen de medida.

La sensibilidad en el punto central de acuerdo a la ecuación (8.13) es:

$$S = 0{,}026\ V/K$$

La recta de calibración según (8.14) resulta:

$$v_{oLineal}(T) = v_o(T_C) + S(T - T_C) = 1{,}06 + 0{,}026(T\text{-}313)$$

Vamos a representar la respuesta real y la linealizada en el margen de medida, figura 8.7a:

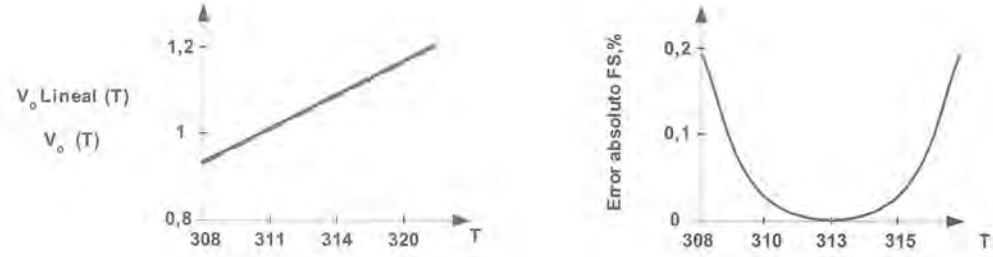


Figura 8.7. (a) Error de *offset* y (b) error de linealidad.

Observamos que la respuesta real y la linealizada son, a la vista de la figura, bastante similares.

El error absoluto (figura 8.7b) se calcula como:

$$e_{absoluto}\ FS\ \% = (v_o - v_{oLineal})/[v_{olineal}(318) - v_{oLineal}(308)] \times 100 \quad (318\ K \equiv 45°C,\ 308\ K \equiv 35°C)$$

Los mayores errores se producen en los extremos como se refleja en la figura 8.7b; no obstante, ambos se mantienen inferiores a 0,2% del FS, es decir, se mantiene el error por debajo de $0{,}2 \cdot 10°C/100 = 0{,}02°C$. El error total es: $0{,}02 + 0{,}01 = 0{,}03°C$ (suma del error de no linealidad y error por autocalentamiento).

El acondicionamiento consistirá en restar el *offset* y amplificar la señal para adaptarla al margen de entrada del convertidor A/D.

El *offset* se calcula como: $v_{oLineal}(308) = 0{,}93$.

La ganancia será: $G = 5/[v_{oLineal}(318) - v_{oLineal}(308)] = 19{,}13$.

El circuito propuesto será:

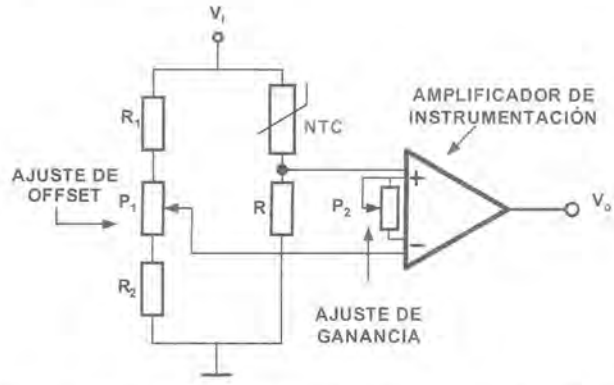


Figura 8.8. Circuito de acondicionamiento para la medida de temperatura.

b) Linealización mediante paralelizado

En ciertas ocasiones se requiere el uso del termistor como un dispositivo linealizado, por ejemplo, en la compensación de dispositivos con coeficientes de temperatura positivos. La forma más sencilla de construir una red de este tipo es mediante una resistencia en paralelo con el termistor, figura 8.9a. La curva R-T del circuito paralelo se muestra en la figura 8.9b.

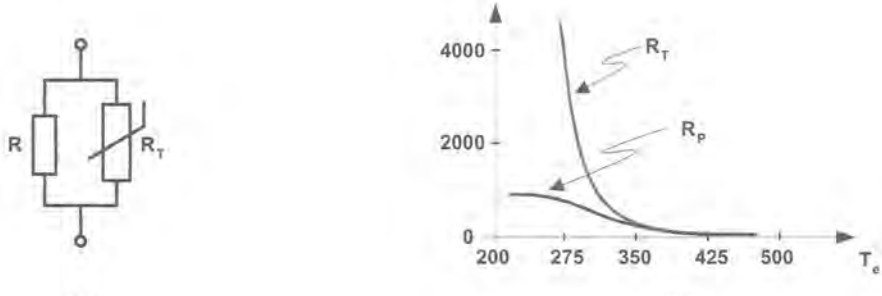


Figura 8.9. (a) circuito de linealización mediante paralelizado; (b) curva del termistor, R_T, y curva con resistencia en paralelo $R_p(T)$. Se observa que se reduce la no linealidad aunque disminuye la sensibilidad.

La resistencia equivalente del conjunto será:

$$R_p(T) = \frac{RR_T}{R + R_T} \tag{8.15}$$

El objetivo, como antes, está en encontrar un valor de R que lleve a unas condiciones óptimas de linealidad. Podemos aplicar el método del punto de inflexión resultando un valor para la resistencia paralelo, R, como el expresado en (8.10).

La sensibilidad del circuito paralelo, $S_p(T)$, puede calcularse a partir de la ecuación (8.15) como:

$$S_p(T) = \frac{dR_p(T)}{dT} = -\frac{B}{T^2} R_T \frac{1}{\left(1 + \frac{R_T}{R}\right)^2} \tag{8.16}$$

La sensibilidad en el punto central de medida, T_c resulta:

$$S_p(T_c) = \frac{dR_p(T)}{dT}\Big|_{T=T_c} = -\frac{B}{T_c^2} R_{Tc} \frac{1}{\left(1 + \frac{R_{Tc}}{R}\right)^2} \tag{8.17}$$

sustituyendo R por la ecuación (8.10) resulta:

$$S_p(T_c) = \frac{dR_p(T)}{dT}\Big|_{T=T_c} = -\frac{1}{4} R_{Tc} \frac{(B - 2T_c)^2}{BT_c^2} \tag{8.18}$$

que puede utilizarse para determinar el valor del termistor necesario para obtener una sensibilidad dada.

La fuente de excitación en los circuitos de linealización mediante paralelizado suele ser una fuente de corriente (figura 8.10).

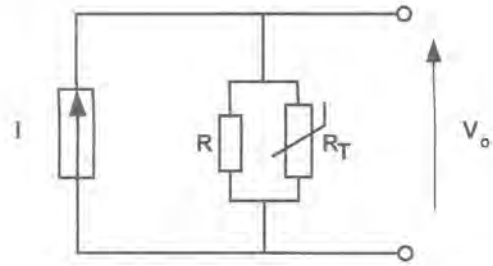


Figura 8.10. Circuito de linealización mediante paralelizado excitado por una fuente de corriente.

8.2.5 OTRAS APLICACIONES DE LAS NTCs

Además de la medida de temperatura, la NTC tiene otras aplicaciones: compensación de dispositivos con coeficiente de temperatura positivo, la detección de nivel de líquidos, medida de velocidad de fluidos, de composición de gases, alarmas de temperatura, protección de sobrecorrientes, etc. A continuación, describiremos brevemente algunas de estas aplicaciones.

Compensación

En estas aplicaciones se utiliza la NTC para contrarrestar las variaciones de dispositivos que tengan un coeficiente de temperatura positivo. Una aplicación típica consiste en compensar los efectos de la temperatu-

ra sobre la bobina de cobre de un galvanómetro. El cobre tiene un coeficiente de temperatura positivo y un aumento de la temperatura provoca un aumento de su resistencia, falseando la medida. Añadiendo una NTC linealizada mediante una resistencia en paralelo puede hacerse que el conjunto tenga un coeficiente de temperatura casi nulo (figura 8.11).

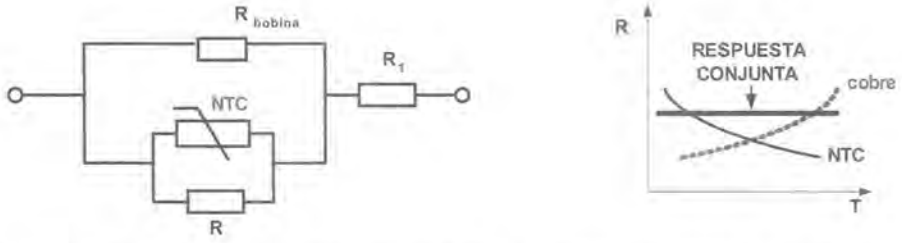


Figura 8.11. Circuito de compensación con una NTC de una resistencia con coeficiente positivo.

Protección contra sobrecorrientes

Esta aplicación está basada en el efecto de autocalentamiento del termistor y su constante de tiempo para generar retardos en la evolución de señales de tensión o corriente. Según el termistor seleccionado, los transitorios pueden ir de milisegundos a varios minutos. La figura 8.12 muestra el circuito de protección para una fuente de alimentación. En el instante inicial los condensadores de filtro están descargados y aparecen como cortocircuitos, produciéndose una elevada corriente de carga que puede llegar a fundir los fusibles de protección. Esta elevada corriente puede evitarse situando un termistor en serie con el dispositivo que se pretenda proteger (figura 8.12). La resistencia inicial de la NTC limita la corriente de carga y, conforme se va autocalentando, su resistencia va disminuyendo hasta alcanzar el valor final estable.

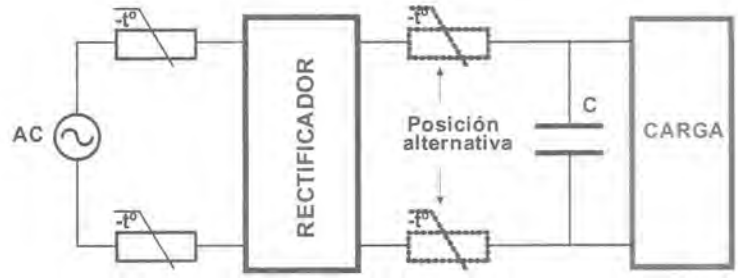


Figura 8.12. Circuito de protección frente a sobrecorrientes utilizando una NTC. Alternativamente podemos situar las NTCs a la salida del puente de diodos.

El termistor utilizado en estas aplicaciones suele ser un tipo especial de muy bajo valor resistivo (cuando ya está caliente) para que la caída de tensión sea despreciable, y con alto coeficiente de disipación. La tabla 8.1 muestra algunos ejemplos.

Tipo [8.10]	R @ 25°C | Ω]	Máxima corriente en régimen permanente [A]	Resistencia máxima en régimen permanente [Ω]
SCL32 0R530	0.5	30	0,01
SCL15 5R006	5	6	0,1
SCL15 22102	220	2	0,8

Tabla 8.1. Termistores de protección de sobrecorrientes, [8.10].

EJERCICIO 8.3

Se pretende proteger los fusibles de 15 A de una fuente en el momento de encendido. La tensión de alimentación proviene de una red de 220 Vrms. Calcule la resistencia mínima que debe tener la NTC.

SOLUCIÓN:

Por seguridad, vamos a limitar la corriente inicial a menos de 10 A. Con una tensión de 220 Vrms resulta una resistencia mínima: de valor $V_{pico}/I_{máx.} \rightarrow 311$ V/10 A = 31 Ω; como las tolerancias en este tipo de NTC rondan el 15% podemos elegir una NTC de 38 Ω @ 25°C.

Por otro lado, tenemos que elegir una NTC cuya corriente en régimen permanente soporte la que se establecerá en nuestro circuito y considerar la energía que es capaz de absorber la NTC para que no falle el sistema. En todas estas aplicaciones deben prevenirse los valores de tensión, corriente o potencia que puedan dañar el termistor.

Medida de la velocidad de fluidos, de niveles de líquidos o análisis de gases

Estas aplicaciones se basan en el cambio en la constante de disipación del termistor (δ). En la zona de autocalentamiento, el termistor es sensible a cualquier acción que altere el ritmo de disipación de calor, como un aumento en la velocidad del fluido que lo circunda o la presencia de un líquido que lo cubra (detección de niveles). En estas situaciones el punto de trabajo del termistor se modifica pasando a un nuevo punto relacionado con la magnitud de la perturbación.

EJERCICIO 8.4

En la figura 8.13a se muestra un depósito de aceite cuyo nivel máximo se pretende detectar con ayuda de una NTC, utilizando para ello el circuito de la figura 8.13b. Calcule el punto de trabajo cuando la NTC pasa de estar al aire con un δ = 16 mW/°C a estar sumergida en aceite con δ = 26 mW/°C. Datos: B = 3290 K, R_0 = 1000 Ω @ 25°C, $P_{máx.}$ = 2 W. Considere que el aceite está a la misma temperatura que el aire, 25°C. ¿Qué sucede si cambia la temperatura ambiente?

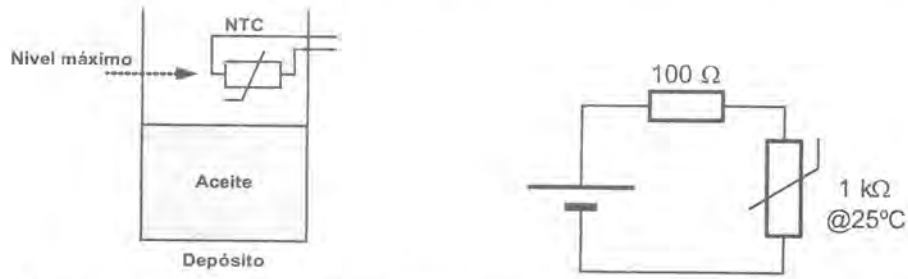


Figura 8.13. (a) Depósito de aceite; (b) circuito de detección del nivel con un termistor.

SOLUCIÓN:

La ecuación (8.7) define la característica tensión-corriente del termistor en condiciones estables de temperatura. Para representarla pueden utilizarse las siguientes ecuaciones:

$$I_T = \sqrt{\frac{\delta(T - T_a)}{R_T}} \quad \text{y} \quad V_T = \sqrt{\delta(T - T_a)R_T} \tag{8.19}$$

donde I_T y V_T son la tensión y la corriente por el termistor en condiciones estacionarias, T_a es la temperatura ambiente y R_T es la resistencia del termistor a la temperatura T.

El punto de trabajo se determina a través de la intersección entre la característica tensión-corriente del termistor y la recta de carga (punto A de la figura 8.14). Cuando la NTC se cubre con el aceite se modifica la constante de disipación, pasando a $\delta = 26$ mW/°C; esta circunstancia hace que se modifique la característica V-I obteniéndose un nuevo punto de trabajo, B. El cambio se produce paulatinamente (según la constante de tiempo) desplazándose la curva V-I sobre los puntos de la recta de carga fija.

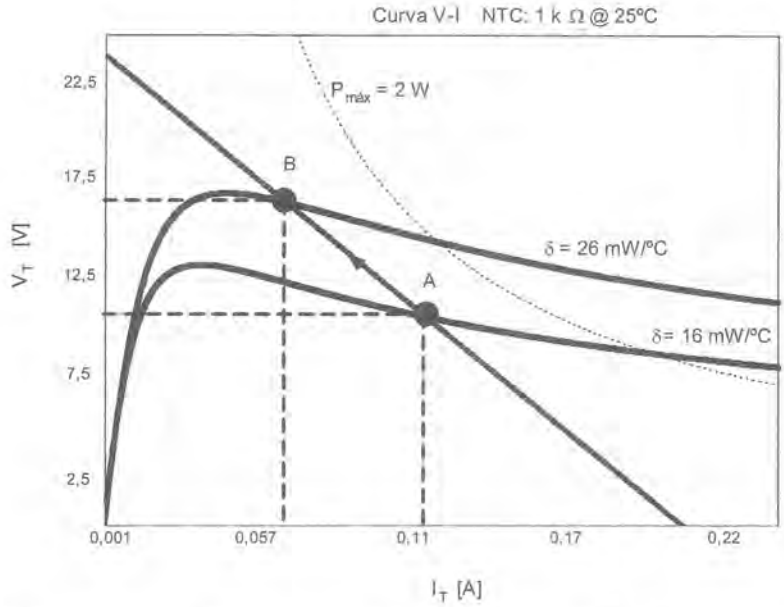


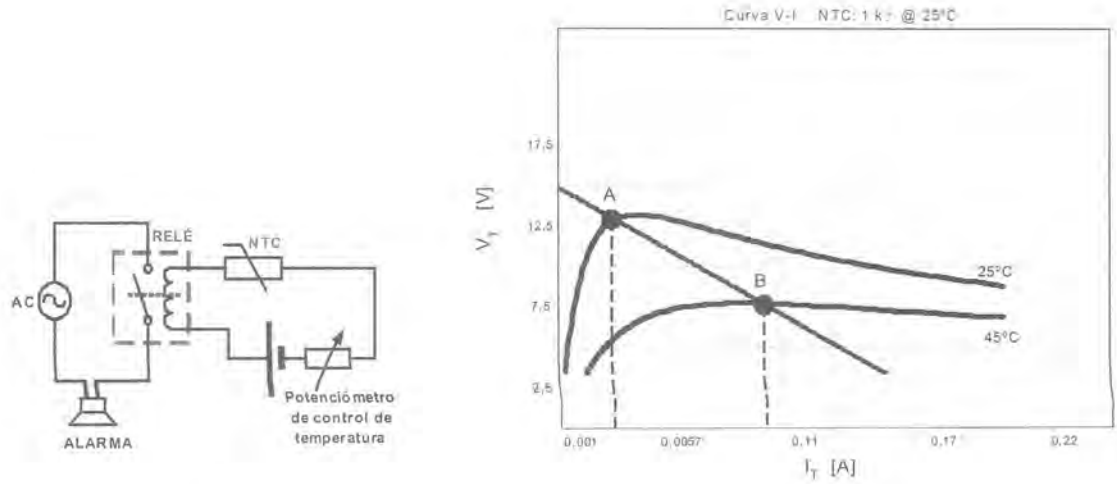
Figura 8.14. Variación de la característica V-I para cambios en la constante de disipación.

El valor de tensión obtenido en la NTC en los puntos A y B puede aprovecharse para discernir en que situación nos encontramos.

Si se modifica la temperatura ambiente, la característica V-I del termistor se modifica lo que puede llevar a errores en la detección.

Alarmas de temperatura

En esta aplicación el punto de trabajo del circuito se ve modificado por cambios en la temperatura ambiente. En la figura 8.15a, se muestra una NTC en serie con la bobina de un relé. Cuando la temperatura ambiente supera un valor preestablecido considerado como el punto de alarma de temperatura, la resistencia de la NTC se reduce, aumentando la corriente lo suficiente como para hacer actuar al relé. En la figura 8.15b se muestran gráficamente los cambios en la característica V-I del termistor ante cambios en la temperatura ambiente (aplicando (8.19)). El punto A es el punto de trabajo del circuito a 25°C y la corriente es insuficiente para activar el relé. A 42°C, punto B, la corriente se incrementa lo suficiente como para activar el relé.



(a) (b)

Figura 8.15. (a) Circuito de alarma de temperatura; (b) al incrementarse la temperatura de 25°C a 45°C la característica V-I desciende cortando a la recta de carga en el punto B. En este punto de trabajo la corriente es suficiente como para activar el relé.

8.3 Termistores de coeficiente de temperatura positivo (PTCs)

8.3.1 PRINCIPIO FÍSICO DE FUNCIONAMIENTO

Las PTCs son termistores con coeficiente de temperatura positivo. Presentan la propiedad de experimentar un cambio brusco en su valor resistivo cuando la temperatura supera un valor crítico característico del material. Están fabricadas con materiales cerámicos policristalinos dopados con impurezas. Se utilizan fundamentalmente compuestos de bario, plomo y titanio con aditivos tales como ytrio, manganeso y tántalo. Su forma más común es un disco con las superficies de contacto metalizadas.

8.3.2 CARACTERÍSTICAS GENERALES

La figura 8.16 muestra la característica resistencia-temperatura de una PTC de conmutación típica. A temperaturas inferiores a T_s, la PTC manifiesta un comportamiento similar al de los semiconductores, con un coeficiente de temperatura negativo (entre 0 y 1 %/°C).

Cuando la temperatura se aproxima a la señalada como T_s, denominada temperatura de Curie, la resistencia del dispositivo se incrementa rápidamente. El coeficiente de temperatura se supone constante en esta zona y puede variar entre 15%/°C a 50%/°C según la PTC (es el valor que suele proporcionarse en las hojas de datos). En la figura 8.16 se observa que, si la temperatura sigue aumentando, el coeficiente se hace de nuevo negativo, aunque las PTCs no suelen utilizarse en esta región. La temperatura T_s puede modificarse

alterando los materiales y el nivel de dopantes. En la mayoría de las PTCs, T_s se sitúa entre 50°C y 160°C, aunque podemos encontrarnos con temperaturas de Curie de 0°C o 300°C.

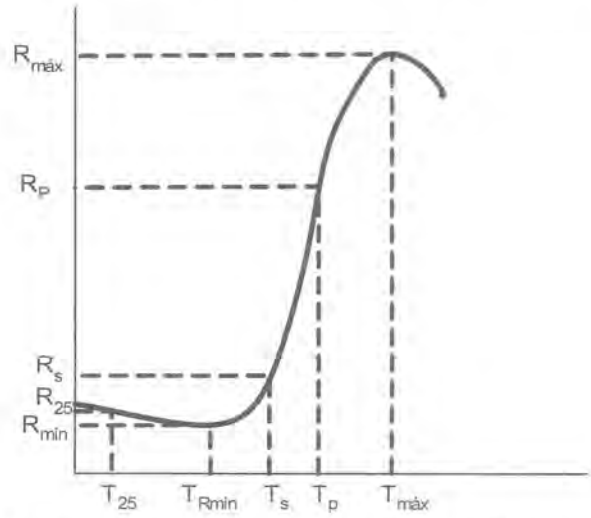


Figura 8.16. Curva resistencia temperatura de una PTC de conmutación.

Las características eléctricas de las PTCs se describen utilizando un conjunto de parámetros ya que no se hace uso de un modelo basado en una ecuación como ocurría con las NTCs. Los parámetros fundamentales son:

- **Resistencia nominal, (R_{25}):** resistencia de la PTC a la temperatura de referencia de 25°C. Esta resistencia se mide en condiciones de potencia nula, de manera que no se produzca autocalentamiento. Sirve como indicación de la resistencia nominal del dispositivo.
- **Resistencia mínima $R_{mín}$:** mínimo valor resistivo que puede adoptar el termistor, asociado a la tempertura $T_{Rmín}$. A partir de este punto, el coeficiente de temperatura empieza a hacerse positivo. Este valor se toma como resistencia base a partir de la cual se calcula la temperatura de conmutación. Normalmente su valor está próximo a R_{25}.
- **Temperatura de conmutación, T_s:** se define como la temperatura en la que la resistencia de la PTC es el doble de la resistencia mínima, R_{min} y se considera el punto donde que se inicia el aumento acusado de resistencia. Esta temperatura puede variar entre –30°C y 270°C según el material de la PTC y presenta cierta tolerancia (±5°C).
- **Resistencia de conmutación, R_s:** resistencia de la PTC a T_s. Se define como el doble de la resistencia mínima ($R_s = 2R_{min}$).
- **Resistencia de transición PTC:** resistencia en el punto de máxima pendiente que se alcanza a la temperatura T_P.
- **Resistencia máxima, $R_{máx}$:** resistencia máxima que se alcanza a la temperatura $T_{máx}$.
- **Temperatura máxima admisible, $T_{máx}$:** temperatura máxima que puede soportar el dispositivo en su superficie sin deteriorarse por causa eléctrica o térmica.
- **Tensión máxima de funcionamiento, $V_{máx}$:** representa la máxima tensión en régimen permanente que puede aplicarse al termistor. Este valor suele incluir un cierto margen de seguridad pero no debe sobrepasarse para no dañar el dispositivo.

Característica tensión-corriente

La característica tensión-corriente define la relación entre estos dos parámetros en condiciones de equilibrio térmico:

$$V_T = \sqrt{\delta(T - T_a)R_T} \tag{8.20}$$

donde R_T representa, en este caso, el valor resistivo de la PTC a la temperatura T. Como no suele proporcionarse una ecuación para modelar la relación R-T vamos a realizar una idealización que consiste en representarla por R_{min} (para $T < T_s$) y por $R_{máx}$ (para $T > T_s$); donde T_s es la temperatura de conmutación (figura 8.17).

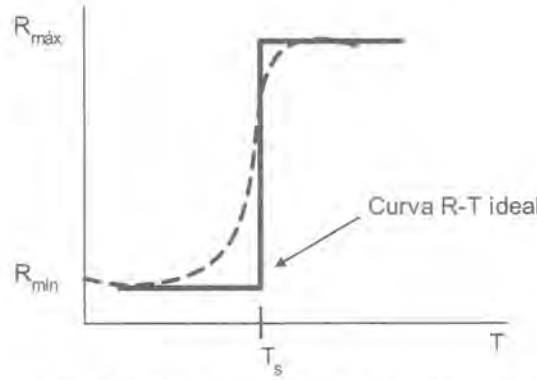


Figura 8.17. Característica R-T ideal de una PTC.

La representación de la caracteristica V-I será, por tanto, una recta de pendiente $1/R_{min}$, representada por el tramo OA de la figura 8.18, hasta alcanzar la tensión crítica V_s, a partir de la cual se produce la conmutación. V_s puede calcularse haciendo $T = T_s$ en (8.20):

$$V_s = \sqrt{\delta(T_s - T_a)R_{min}} \quad \text{con} \quad I_{ms} = \sqrt{\delta(T_s - T_a)/R_{min}} \tag{8.21}$$

La corriente I_{ms} será la mínima corriente necesaria para conmutar la PTC a su estado de alta resistencia. A partir de este punto se produce la transición hacia el valor resistivo máximo en la PTC, $R_{máx}$. Si suponemos constante la temperatura T_s durante este tránsito, la potencia disipada también será constante. La representación en la característica V-I se corresponde con una hipérbola equilátera (potencia = vi = cte) señalado como el tramo AB, hasta alcanzar la tensión V_m, momento en que se alcanza la resistencia máxima $R_{máx}$. V_m puede calcularse como:

$$V_m = \sqrt{\delta(T_s - T_a)R_{máx}} \tag{8.22}$$

A partir de aquí, la resistencia en la PTC se mantiene constante, $R_{máx}$, y se representa por el tramo OB de pendiente $1/R_{máx}$. La corriente en esta zona es I_r que es una corriente residual de bajo valor debido al alto valor resistivo de la PTC en este punto. La tensión $V_{máx}$ señala la máxima tensión aplicable a la PTC. Si se supera, se puede producir la ruptura en tensión, pudiendo deteriorar el dispositivo (figura 8.18). También se dibuja la curva para una temperatura ambiente T_b. La gráfica se desplaza hacia abajo A'B' pero se mantienen las rectas $1/R_{min}$ y $1/R_{máx}$.

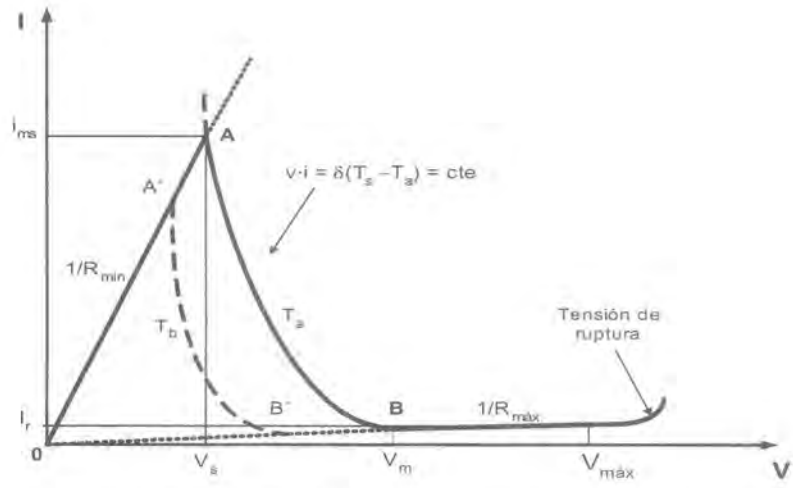


Figura 8.18. (a) Característica V-I de una PTC a temperatura ambiente constante. Si la temperatura ambiente aumenta, $T_b > T_a$, la curva AB se desplaza hacia abajo, A'B' manteniéndose las rectas $1/R_{mín}$ y $1/R_{máx}$ [8.12].

EJERCICIO 8.5

Indique cómo afectan los cambios en la temperatura ambiente (de 25ºC a 50ºC) y en la constante de disipación (de 12 a 6 mW/ºC) a la característica V-I. ¿Qué debería cumplir un circuito para que a 50ºC se produzca la conmutación en la PTC?

SOLUCIÓN:

Al aumentar la temperatura ambiente la característica V-I se desplaza hacia abajo (8.19a). Si aumenta la constante de disipación, la gráfica se desplaza hacia arriba (8.19b). Como ocurría en las NTCs, esta circuns-

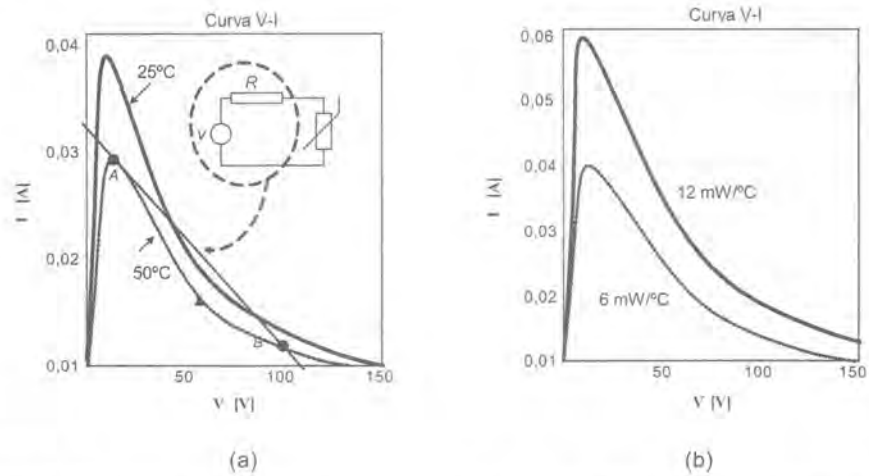


Figura 8.19. Alteraciones en la característica V-I de una PTC; (a) para cambios en la temperatura ambiente y (b) cambios en la constante de disipación.

tancia puede utilizarse para medir variables que alteren estos parámetros y que, por tanto, lleven a cambios en el punto de trabajo del circuito. Para que a 50°C se entre en la zona de alta resistencia la recta de carga del circuito que se conecta a la PTC debe diseñarse como se muestra en la figura 8.19a. El punto A sobre la característica V-I a 50°C es inestable desplazándose hasta el punto B que es estable.

8.3.3 APLICACIONES

La mayor parte de las aplicaciones de las PTCs se derivan de su característica V-I en condiciones estables de autocalentamiento o de su característica corriente-tiempo en condiciones dinámicas de autocalentamiento o de una combinación de ambas. Existen menos aplicaciones basadas en la característica R-T (medida de temperatura), aunque también pueden utilizarse para la supervisión de temperatura de sistemas y máquinas. Entre las aplicaciones (basadas en el autocalentamiento) podemos destacar: calentadores auto-regulados, protección de sobrecorrientes (fusibles reseteables), supresión de arcos, detección de niveles de líquidos, arranque de motores, retardos de tiempo, etc.

Calentadores auto-regulados

Estas aplicaciones utilizan la PTC como un calentador auto-regulado que mantiene una temperatura constante para un margen amplio de tensiones de alimentación y condiciones de disipación. Su capacidad de autoregulación permite la eliminación del termostato u otro circuito de control lo que abarata costes y simplifica el diseño. Por ejemplo, se utilizan PTCs en diversos actuadores de tipo térmico como los conmutadores bimetálicos o en el calentamiento de una cera para fundirla, o de un líquido para vaporizarlo o para controlar su viscosidad (se utiliza en las impresoras para controlar la fluidez de la tinta). También se utilizan como generadores de calor en diversos aparatos domésticos (rizadores de pelo), etc. Su funcionamiento como calentador es bastante simple: cuando se aplica una tensión a una PTC comienza a circular una corriente. Si la tensión aplicada es suficiente, la PTC se autocalienta hasta que alcanza su región de resistencia positiva (PTC). En este punto se alcanza un equilibrio cuando toda la potencia suministrada se disipa al ambiente manteniendo la PTC una temperatura determinada. El efecto de autorregulación se manifiesta por el hecho de alcanzar el equilibrio en la región de resistencia positiva. Si, por ejemplo, la temperatura se redujese (por cambio en la temperatura ambiente o por otro motivo), la resistencia disminuiría haciendo circular más corriente y, si la temperatura aumentase, se produciría el efecto contrario. Incluso si se modifica la tensión de alimentación, el mecanismo de autorregulación sigue siendo efectivo (si se incrementa la tensión y se proporciona más potencia, la temperatura se incrementa, la resistencia aumenta, y se reduce la corriente). En general, la potencia consumida por una PTC es prácticamente independiente de la tensión aplicada en un margen bastante amplio.

Cuando se selecciona una PTC para este tipo de aplicación deben considerarse varios factores aunque el parámetro más importante es la temperatura de conmutación T_s ya que la temperatura máxima que alcanza la PTC es unos grados superior a esta temperatura.

Protección frente a sobrecorrientes. Fusibles reseteables

El incremento tan brusco de la resistencia en la PTC por encima de la temperatura de transición las hace ideales para la protección de circuitos frente a sobrecorrientes. El circuito debe diseñarse de manera que para las corrientes por debajo del límite impuesto, la potencia disipada en el termistor no sea suficiente para llevarlo, a través del autocalentamiento, a su temperatura de transición T_s. Si la corriente supera un límite debido a un fallo, por la PTC circulará más corriente y el efecto de autocalentamiento hará que se alcance la temperatura de transición y aumente bruscamente la resistencia, limitando la corriente en el circuito. Tan pronto como se restablezcan las condiciones normales en el circuito, el sistema volverá a su estado normal,

es decir, funciona como una especie de "fusible reseteable". En este tipo de aplicación la PTC se sitúa en serie con el componente que se pretenda proteger, como muestra la figura 8.20.

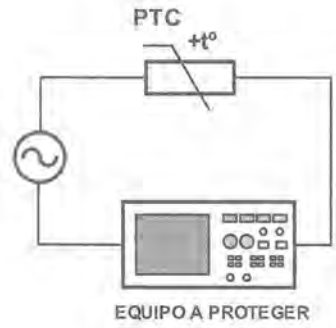


Figura 8.20. Circuito de protección frente a sobrecorrientes.

Las PTCs como elementos de conmutación con retardo

Estas aplicaciones están basadas en el funcionamiento dinámico de una PTC autocalentada (carácterística corriente-tiempo). La corriente pasa durante un tiempo hasta que el efecto de autocalentamiento hace que la resistencia aumente bruscamente y reduzca la corriente. Este tiempo de conmutación depende de varios factores: tamaño de la PTC, resistencia, temperatura ambiente, etc. y puede variarse entre ciertos límites, pudiendo utilizarse para realizar distintos retardos. Las aplicaciones típicas incluyen el arranque de motores, los procesos de desmagnetización, el arranque de lámparas de descarga [8.13], etc. Por ejemplo, los circuitos de la figura 8.21 permiten generar retardos en la conexión y en la desconexión de un relé. En el primer caso, se sitúa una PTC en paralelo con la bobina del relé (figura 8.21a). En el momento de conexión de la alimentación (se cierra el interruptor I) toda la corriente circula por la PTC debido a su bajo valor resistivo inicial, permaneciendo abierto el interruptor del relé. Conforme se calienta, llega un momento en que se alcanza la temperatura de conmutación; en ese instante, toda la corriente pasa a circular por la bobina del relé cerrándose su contacto. En 8.21b se muestra la configuración para generar un retardo en la desconexión del relé. En el momento inicial el contacto del relé está cerrado hasta que la PTC conmute y limite la corriente haciendo que se abra.

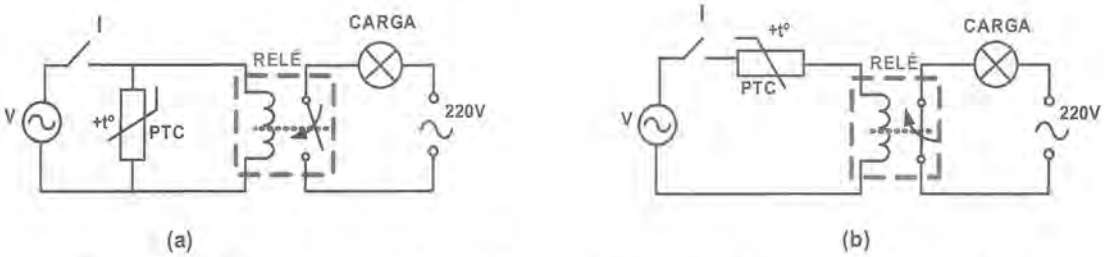


Figura 8.21. Circuitos con PTCs para generar (a) retardos de tiempo en la conexión y (b) retardos en la desconexión.

Supresión de arcos

Esta aplicación permite eliminar la generación de sobretensiones que se producen en los interruptores que conmutan cargas fuertemente inductivas[1] (figura 8.22).

[1] Cuando se corta la corriente por un circuito fuertemente inductivo, la tensión originada, proporcional a la derivada de corriente, resulta muy elevada pudiendo producir arcos en el dispositivo de corte.

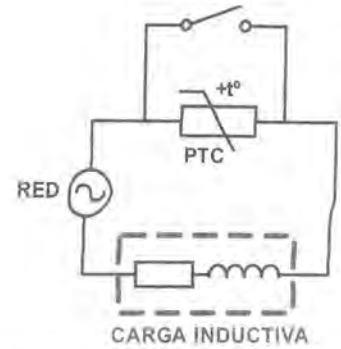


Figura 8.22. Protección frente a sobrecorrientes.

Situando una PTC en paralelo con el interruptor se evita la sobretensión ya que la PTC permite la recirculación de la corriente hasta que conmuta a su zona de alta resistencia.

Detección de nivel y de flujo en líquidos o gases

Estas aplicaciones se basan en el cambio en el coeficiente de disipación y su efecto en la característica V-I. Un incremento en el coeficiente de disipación, situando el dispositivo en un líquido o en una corriente de aire, hará que se modifique su punto de trabajo de forma similar a como vimos en la NTC.

8.4 Fotorresistencias (LDRs)

8.4.1 PRINCIPIO DE FUNCIONAMIENTO

Las **fotorresistencias** (**LDRs,** *Light Dependent Resistors,* resistencias dependientes de la luz) son sensores resistivos basados en semiconductores empleados para la medida y detección de radiación electromagnética.

Una LDR está constituida por un bloque de material semiconductor sobre el que puede incidir la radicación y dos electrodos metálicos en los extremos (figura 8.23).

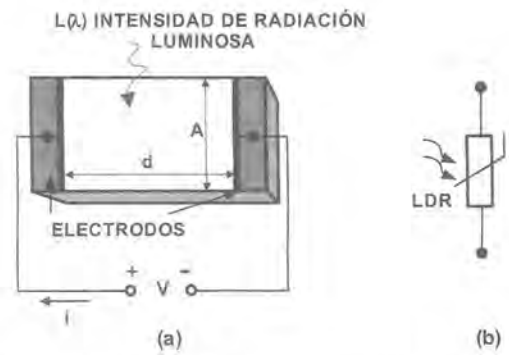


Figura 8.23. (a) Estructura de una LDR fabricada con material semiconductor fotosensible; (b) símbolo eléctrico.

La conductividad en los semiconductores depende del número de portadores de carga capaces de moverse. En un material semiconductor, la anchura de la banda prohibida o diferencia energética entre las bandas de valencia y conducción, E_G, es intermedia entre los aislantes y los conductores. A bajas temperaturas, los semiconductores se comportan como aislantes porque casi todos los electrones se encuentran en la banda de valencia unidos a sus átomos. La aportación de energía, bien mediante un aumento de temperatura (como vimos hasta ahora) o mediante radiación electromagnética, puede modificar esta situación.

En el caso de radiación electromagnética, si un fotón de frecuencia υ y energía E_p ($E_p = h \cdot \upsilon$, donde h es la constante de Planck) incide sobre un electrón, puede ocurrir que la energía absorbida por éste sea superior a la energía de la banda prohibida E_G, en cuyo caso, el electrón se sitúa en un nivel energético superior dentro de la banda de conducción, dejando, a su vez, un hueco en la banda de valencia. La energía necesaria para llevar a cabo este proceso obliga a que $E_p > E_G$. E_G es una constante que depende del tipo de semiconductor; por tanto, existe una frecuencia mínima, υ_g, a partir de la cual se cumple la desigualdad anterior $\upsilon_g > E_G/h$. Expresado en términos de longitud de onda λ, resulta:

$$\lambda_g < hc/E_G \qquad (8.23)$$

siendo *c* la velocidad de la luz en el vacío. Esta expresión establece una longitud de onda máxima, λ_g, para conseguir la transición del electrón. La presencia de electrones en la banda de conducción y de huecos en la de valencia inducidos por fotones de suficiente energía da lugar a un incremento en la conductividad del semiconductor. Este fenómeno se denomina **fotoconductividad** y es el fundamento de los **sensores fotoconductivos**.

EJERCICIO 8.6

Calcule la máxima longitud de onda que permite la transición de electrones en un material semiconductor como el CdS, donde $E_G = 2{,}4$ eV y en CdSe con $E_G = 1{,}8$ eV.

SOLUCIÓN:

Aplicando (8.23) y teniendo en cuenta: 1 eV = $1{,}602\ 10^{-19}$ J ; h = $6.626\ 10^{-34}$ Js ; c = $3\ 10^8$ m/s, resulta: $\lambda_g < 6{,}626 \cdot 10^{-34}$ Js $\cdot\ 3 \cdot 10^8$ m/s / $2{,}4(1{,}602 \cdot 10^{-19}$ J) $\approx$ 517 nm.

Luego el CdS detecta longitudes de onda más cortas que 517 nm.

En el caso del CdSe se obtiene $\lambda_g < 689$ nm.

La sensibilidad del material de la LDR puede definirse como el número de electrones generados por un fotón y se calcula como $s = V\mu_n \tau_n/d^2$ [8.5] donde d es la distancia entre los electrodos, V es la tensión aplicada, μ_n la movilidad de los electrones y τ_n es el tiempo de vida del electrón antes de que se recombine.

EJERCICIO 8.7

Calcule la sensibilidad de un material fotorresistivo al que se aplica una tensión de 1 V conociendo que: $\mu_n = 290\ cm^2/Vs$, $\tau_n = 10^{-3}s$, $d = 0{,}1$ mm.

SOLUCIÓN:

Aplicando la fórmula $s = V\mu_n \tau_n/d^2$ resulta: s = 2900, la sensibilidad puede ser muy alta, un simple fotón puede liberar 2900 electrones para la conducción.

Tipos y construcción

Una LDR típica consiste en una fina capa semiconductora dispuesta sobre un sustrato cerámico o plástico. La película semiconductora describe una pista en zig-zag con contactos metálicos en los extremos (figura 8.24). La forma de la película sensitiva tiene por objeto maximizar la superficie de exposición y al mismo tiempo mantener un espacio reducido entre los electrodos para aumentar la sensibilidad.

Entre los materiales semiconductores más usados para realizar las LDRs se encuentran el sulfuro de cadmio (CdS) y el seleniuro de Cadmio (CdSe). Estos dos materiales son muy utilizados, sobre todo en los sensores fotoconductivos de bajo coste; sin embargo, se caracterizan por una respuesta relativamente lenta (su constante de tiempo es del orden de 50 ms). La tensión máxima que pueden soportar puede llegar hasta 600 V y hay modelos capaces de disipar más de un vatio. Otros materiales pueden ser: PbS, InSb, $Hg_xCd_{1-x}Te$, y germanio dopado con cinc y boro. Cada compuesto presenta un espectro de detección característico.

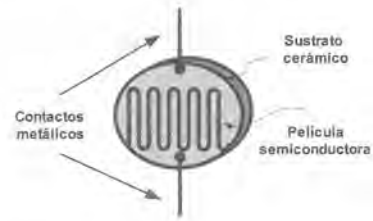


Figura 8.24. LDR con una pista en zig-zag de material resistivo.

8.4.2 MODELO DE LA LDR

La relación entre la resistencia de la LDR y la iluminación[7] L puede modelarse a partir de la ecuación:

$$R_L = R_0 \left(\frac{L_0}{L} \right)^{\alpha} \tag{8.24}$$

donde L es la iluminación (en lux), α es una constante que depende del material (entre 0,7 y 1,5) y R_L y R_0 son las resistencias a los niveles de luz L y L_0 respectivamente. La expresión (8.24) se representa en la figura 8.25:

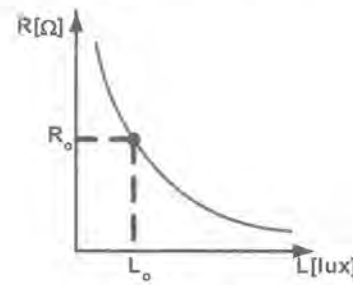


Figura 8.25. Relación resistencia-iluminación en una LDR. La diferencia entre la resistencia con niveles altos de luz y en la oscuridad es muy alta (la relación es mayor de 10.000) y la sensibilidad para bajos niveles de iluminación es muy acusada.

[7] Véase Apéndice 3 en el CD adjunto para conocer las unidades fotométricas.

La tabla 8.2 muestra los valores más habituales de los parámetros del modelo junto a otros datos característicos.

Parámetro	Valores típicos
R_0	2 K-200 K @ 10 lux
α	0,7-1,5
Tensión máxima	100 V-150 V
Potencia máxima	50 mW-1 W
Tiempo de respuesta	>10 ms a varios s

Tabla 8.2. Valores típicos de los parámetros de las LDRs.

8.4.3 APLICACIONES DE LA LDR

Las aplicaciones de las LDRs se centran en la medida de luz (fotometría) y en la detección de cambios de luz. Las fotorresistencias son sensores bastante sensibles, sobre todo con bajos niveles de iluminación, aunque en esta zona su dependencia con la temperatura se hace muy acusada. Otro inconveniente es su lentitud de respuesta ya que presentan constantes de tiempo del orden de segundos lo que limita su uso a aplicaciones de baja frecuencia.

Acondicionamiento de la LDR

El acondicionamiento suele realizarse mediante un divisor de tensión, siendo aplicables las consideraciones realizadas para las NTCs (figura 8.26).

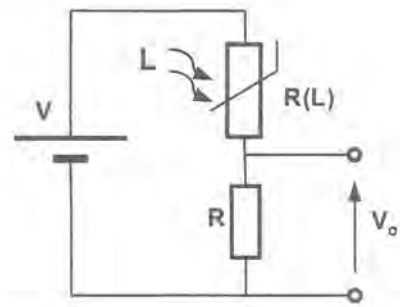


Figura 8.26. Acondicionamiento de una LDR mediante un divisor resistivo.

Así, en aplicaciones de medida donde se busca una respuesta lineal, la obtención del valor de R puede hacerse calculando el valor que lleve a la máxima linealidad en el punto central del intervalo de medida (puede aplicarse el método del punto de inflexión visto anteriormente). Para que exista punto de inflexión debe cumplirse la condición $\alpha > 1$ por lo que deberá tenerse en cuenta a la hora de elegir la LDR en aplicaciones de fotometría.

Para el circuito de la figura 8.26 se obtiene la tensión de salida:

$$V_o = \frac{R}{R + R_o\left(\frac{L_o}{L}\right)^{\alpha}} V \qquad (8.25)$$

Haciendo que el punto de inflexión de la expresión anterior coincida con el centro de nuestro intervalo de medida, se obtiene un valor de R dado por:

$$R = \frac{\alpha - 1}{\alpha + 1} R_{L_C} \tag{8.26}$$

donde R_{Lc} es la resistencia de la LDR en el punto central del intervalo de medida. La sensibilidad en ese punto se calcula como:

$$S(L_C) = \frac{V}{L_c} \frac{\alpha^2 - 1}{4\alpha} \tag{8.27}$$

El valor de V debe limitarse para que no exista autocalentamiento, resultando:

$$V < 2\sqrt{\delta \Delta T R} \tag{8.28}$$

siendo δ el coeficiente de disipación y ΔT el incremento de temperatura máximo permitido por autocalentamiento.

En general, se recomienda que los dispositivos tengan alta resistencia para niveles de iluminación intensos y baja resistencia para niveles débiles [8.12]. También es posible, como en las NTCs, considerar la LDR como un sensor más linealizado conectando una resistencia en paralelo. Como sucedía con las NTCs, se disminuye la sensibilidad aunque el dispositivo es más lineal.

EJERCICIO 8.8

Se desea controlar mediante una LDR una lámpara, de forma que se encienda cuando la luz ambiente descienda por debajo de 20 lux. La lámpara está gobernada por un actuador que deberá recibir una señal superior a 3 V para proceder a su encendido.

Se proporcionan como datos de la LDR: R_o = 10000 Ω @ 10 lux y α = 1.

SOLUCIÓN:

Calculamos el valor resistivo de la LDR a 20 lux a partir de (8.24)

$$R_{20} = 10000(10/20)^1 = 5000\ \Omega$$

Utilizamos el circuito de la figura 8.27, un amplificador inversor con $V_i = -1$ y la LDR en la rama de realimentación.

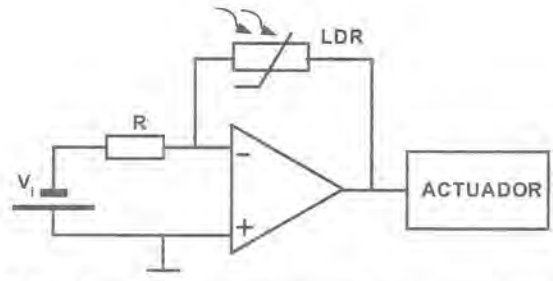


Figura 8.27. Circuito de control de una lámpara con una LDR.

La resistencia de entrada R necesaria para que a la salida se obtengan 3 V a partir de R_{20} es:

$v_o = -[R_{20}/R]\ (-v_i)$, luego $3 = -[5000/R]\ (-1)$, de donde se obtiene R = 1K6

La corriente proporcionada es muy pequeña, así que apenas se modifica el valor calculado de R_{20} por efecto del autocalentamiento. También puede sustituirse R por un potenciómetro para ajustar el umbral de encendido.

Resumen

- Los termistores son sensores de temperatura de tipo resistivo fabricados a partir de materiales semiconductores cuya conductividad se modifica con la temperatura. Se dividen en dos grupos fundamentales: NTC (*Negative Temperature Coefficient)* y PTC (*Positive Temperature Coefficient).*
- Las NTCs se caracterizan por presentar un coeficiente de temperatura negativo, es decir, la resistencia de estos dispositivos disminuye cuando aumenta la temperatura y, además, lo hace de forma no lineal. Presenta como característica más relevante una alta sensibilidad, hasta 10 veces mayor que una RTD. Las NTCs son componentes muy robustos, fiables y económicos. Sus mayores inconvenientes son: su lentitud de respuesta, la presencia de grandes tolerancias en su fabricación (20%), una relativa estabilidad (1%/año) y un campo de medida también algo limitado (–70°C a 300°C). Entre las aplicaciones básicas están la medida de temperatura aunque también se utilizan para compensar dispositivos con coeficiente positivo, medidas de nivel y velocidad de fluidos, protección de sobrecorrientes, etc.
- Las PTCs son termistores con coeficiente de temperatura positivo, es decir, su resistencia aumenta al incrementarse la temperatura. Su característica esencial es que la resistencia aumenta de forma muy brusca a partir de una temperatura determinada.
- Las PTCs se utilizan fundamentalmente como calentadores autorregulados, protección de sobrecorrientes, supresión de arcos, detección de niveles de líquidos, generación de retardos en señales eléctricas, etc.
- Las fotorresistencias (LDRs) son sensores resistivos basados en el cambio en la conductividad que experimenta un semiconductor al incidir sobre él radiación electromagnética. Se utilizan para medir luz (fotometría) y en aplicaciones de detección de presencia y posición.

Problemas propuestos

8.1 Calcule el valor resistivo de una NTC a 26,5°C conociendo que a 25°C su valor es de 10K y α = –4,28%/°C a 25°C.

8.2 Diseñe un sistema de medida de temperatura entre 50 y 100°C con un error inferior a 2°C, utilizando una NTC con los siguientes datos: B = 4600K, R_0= 100K @ 25°C y δ = 10 mW/°C.

8.3 Se pretende medir temperatura entre 25 y 50°C con una precisión de 0,1°C mediante una NTC utilizando un microntrolador para realizar el cálculo de la temperatura a partir del modelo exponencial. En la tabla adjunta se proporciona un conjunto de datos de calibración del termistor. Obtenga los coeficientes del modelo de la NTC por regresión lineal sobre los datos proporcionados. Compare los resultados del modelo obtenido con los datos experimentales para estimar los errores cometidos. Proponga el circuito de medida.

Datos de calibración	
Resistencia medida [Ω]	**Temperatura medida [°C]**
1102	25,214
912,4	30,140
755,7	35,305
637,0	40,200
534,2	45,213
452,3	50,227

8.4 El siguiente circuito convierte la resistencia del termistor a una frecuencia utilizando un multivibrador astable (555). Calcule el periodo de la señal cuadrada obtenida e indique su posible utilidad. Proponga un método para el ajuste de la pendiente y el *offset*. ¿Qué problemas plantea esta solución?

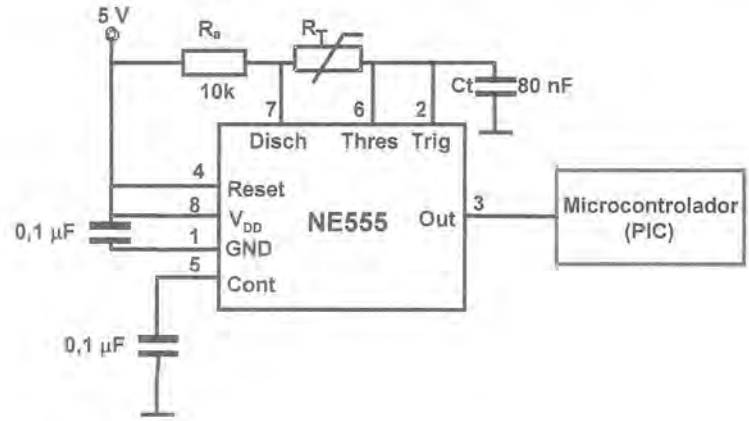


Figura 8.28. Circuito de medida de temperatura basado en un multivibrador.

8.5 Diseñe un sistema de alarma de temperatura utilizando una NTC. Deberá activarse cuando la temperatura ascienda por encima de 100°C con una precisión de 1°C (**Datos:** B = 3600K, R_0 = 100K @ 25°C, δ = 10 mW/°C y alimentación 5 V).

8.6 Proponga un circuito que actúe como un regulador de temperatura para calefacciones eléctricas de manera que el usuario pueda escoger entre 3 temperaturas: 100°C, 120°C y 80°C. La calefacción deberá permanecer encendida hasta alcanzar la temperatura seleccionada en cuyo instante deberá apagarse. Suponga que la calefacción equivale a una carga resistiva de 80 Ω, se alimenta a red (220 VAC) y dispone de un TRIAC como elemento conmutador.

8.7 El siguiente circuito se comporta como una fuente de corriente; analice su funcionamiento y proponga una ecuación para el cálculo de la corriente I_s.

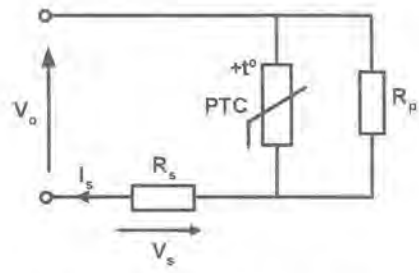


Figura 8.29. Fuente de corriente con una PTC

8.8 Dada una LDR con $\alpha = 1{,}2$ y $R_0 = 20$ kΩ @ 20 lux, diseñe un sistema de medida con la máxima sensibilidad en el margen de 15 a 55 lux. La salida se introduce a un conversor A/D de entrada 0-5 V y 100 kΩ de impedancia de entrada Se pretende que el error por autocalentamiento sea inferior a 0,1°C.

8.9 Se pretende encender de forma automática un sistema de riego cuando la luz descienda por debajo de 100 lux. Deberá proporcionarse una señal de 5 V a un actuador. Diseñe el sistema utilizando una LDR ($R_0 = 20000$ @ 20 lux y $\alpha = 0{,}7$).

9

Otros sensores resistivos

9.1 Sensores de gases de óxidos metálicos semiconductores
9.2 Magnetorresistencias

9.1 Sensores de gases de óxidos metálicos semiconductores

Los sensores de gases resistivos están compuestos de óxidos metálicos, generalmente de estaño (SnO_2). En la figura 9.1a se muestra la estructura de los cristales de SnO_2 que componen el material. Cuando la temperatura es suficientemente elevada y en presencia de oxígeno (figura 9.1b), este es adsorbido por el material al que le transfiere un electrón. La superficie del cristal queda cargada negativamente de forma que se establece una barrera de potencial que dificulta la circulación de corriente aumentando, por tanto, la resistencia.

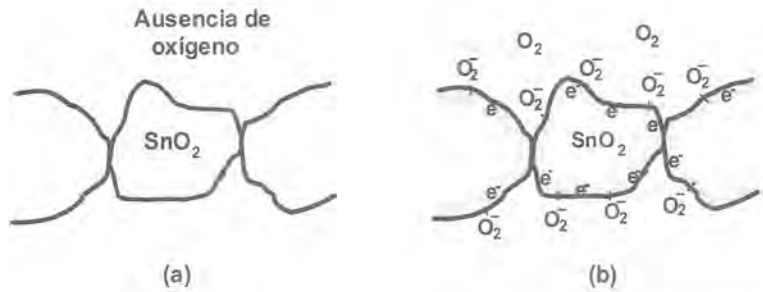


Figura 9.1. (a) Esquema de los cristales que forman un sensor de gases de óxido de estaño; (b) en presencia de oxígeno la superficie de los cristales se carga negativamente.

Se trata, en principio, de un sensor útil para medir la concentración de oxígeno; pero, además, indirectamente se puede medir la concentración de otros gases. Así, en presencia de un gas reductor (por ejemplo,

CO que al reaccionar dará lugar a CO_2), disminuye la concentración de oxígeno adsorbido lo que disminuirá la barrera de potencial y, por tanto, la resistencia. En este caso, el comportamiento sigue una relación exponencial [9.3] (figura 9.2):

$$R_{[C]} = K \cdot [C]^{-a} \tag{9.1}$$

donde K y a son constantes, [C] es la concentración del gas de que se trate y $R_{[C]}$ es la resistencia a esa concentración.

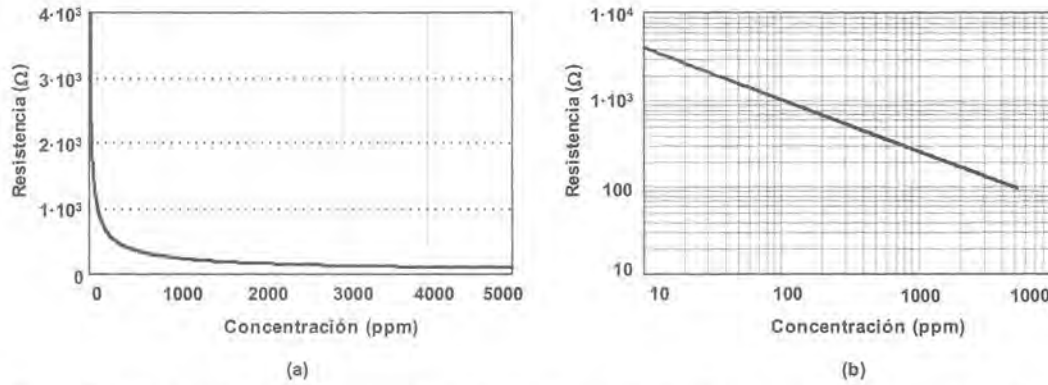


Figura 9.2. Variación de la resistencia con la concentración de hidrógeno en el sensor TGS821 de la firma Figaro; (a) representación en coordenadas lineales; (b) representación en coordenadas logarítmicas.

EJERCICIO 9.1

Las hojas de características del sensor de hidrógeno TGS821 de la firma Figaro indican que la resistencia a una concentración de hidrógeno de 100 ppm (partes por millón) es de 1 kΩ. Además, indican la relación entre las resistencias a 100 ppm y a 1000 ppm según la siguiente expresión:

$$\frac{\log\left(R_{100\ ppm}/R_{1000\ ppm}\right)}{\log(1000\ ppm/100\ ppm)} = 0,6$$

¿Cuál es la curva de calibración según la expresión (9.1)?

SOLUCIÓN:

Según la expresión (9.1):

$$R_{100\ ppm} = K \cdot 100^{-a}$$

$$R_{1000\ ppm} = K \cdot 1000^{-a}$$

Dividendo estas dos expresiones y tomando logaritmos, se obtiene:

$$\frac{\log\left(R_{100\ ppm}/R_{1000\ ppm}\right)}{\log(1000\ ppm/100\ ppm)} = a$$

Luego a = 0,6. Además, el valor de K se puede obtener teniendo en cuenta el valor de la resistencia a 100 ppm. Así:

$$1000 = K \cdot 100^{-0,6}$$

obteniéndose $K = 1,6 \cdot 10^4$. Por lo tanto, la curva de calibración es:

$$R_{[C]} = 1,6 \cdot 10^4 \cdot [C]^{-0,6}$$

donde la concentración está expresada en ppm.

Los sensores presentan cuatro terminales (figura 9.3a), dos corresponden al propio sensor y otros dos, a los electrodos de calentamiento de forma que la temperatura sea la adecuada para que tengan lugar las reacciones. La tensión para alimentar estos electrodos suele ser reducida (5-10 V) y la potencia consumida está alrededor de 1 W.

Uno de los inconvenientes de estos sensores es su escasa selectividad puesto que, en principio, son sensibles a cualquier gas que modifique la concentración de oxigeno. Realmente, puede conseguirse una mayor sensibilidad a un gas concreto en función de la temperatura de los electrodos y del filtro interno que incorporen (figura 9.3b). No obstante, ello no impide que también respondan a otros gases además de al deseado *(figura 9.4)*.

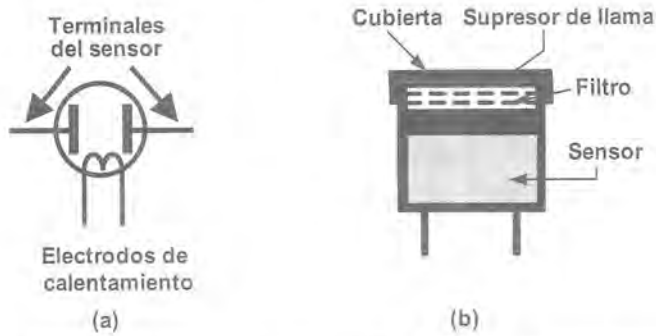


Figura 9.3. (a) Símbolo de un sensor de gases de óxidos metálicos semiconductores; (b) estructura interna.

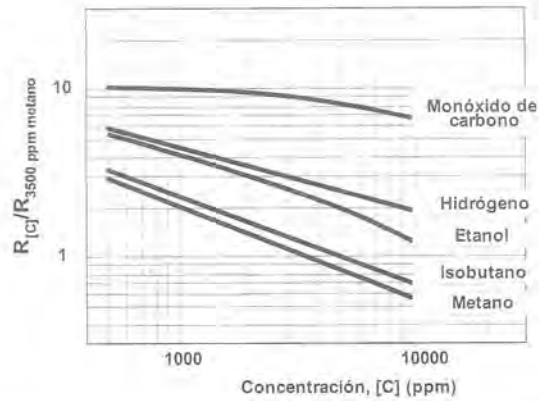


Figura 9.4. Respuesta del sensor de metano TGS842 al propio metano y a otros gases (tomado de [9.4]).

Otros inconvenientes de gran importancia residen en la fuerte dependencia con respecto a la temperatura y a la humedad. Esto hace que sean sensores adecuados para la detección de niveles de alarma más que para la medida de concentraciones concretas.

Con respecto a los circuitos de acondicionamiento se deben de tener en cuenta varias cuestiones:

- Las tolerancias de fabricación son generalmente muy grandes por lo que, normalmente, se necesita algún tipo de calibración.
- Para obtener unos buenos resultados se necesita realizar algún tipo de compensación con respecto a la temperatura y, si fuera necesario, con respecto a la humedad. En la referencia [9.3] se muestra un circuito de acondicionamiento con compensación de la temperatura empleando una NTC.
- Los electrodos de calentamiento tardan algún tiempo en alcanzar la temperatura nominal desde que se alimenta el dispositivo (30-60s). En los instantes iniciales la resistencia es baja (puesto que no se adhiere el oxígeno) y podría ser interpretado como una concentración muy elevada del gas a detectar. Por tanto, la medida durante el tiempo de calentamiento no es válida.

El circuito de acondicionamiento habitual consiste, esencialmente, en una resistencia de carga (figura 9.5). En las aplicaciones consistentes en determinar niveles de alarma basta comparar la tensión de salida con la tensión correspondiente al nivel de alarma. Además, en estas aplicaciones la resistencia de carga se escoge del mismo valor óhmico que la correspondiente al sensor a la concentración de alarma (véase el ejercicio 9.2).

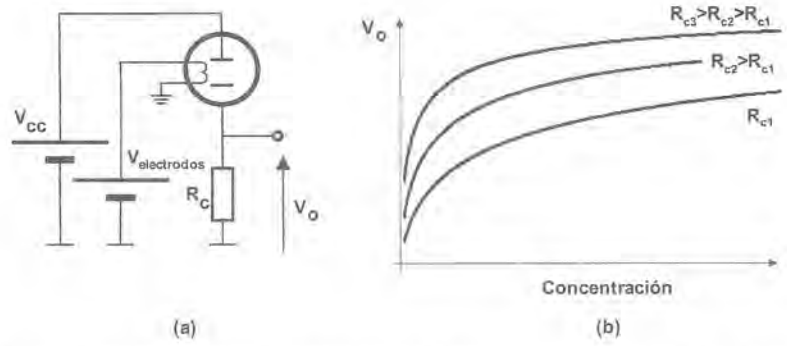


Figura 9.5. (a) Circuito de acondicionamiento típico; (b) tensión de salida V_o en función de la concentración del gas a detectar y de la resistencia de carga, R_c.

EJERCICIO 9.2

Escoja la resistencia de carga para el circuito de acondicionamiento del sensor de hidrógeno mencionado en el ejercicio 9.1, sabiendo que se desea activar una alarma cuando la concentración de hidrógeno supere 100 ppm.

SOLUCIÓN:

Se tratará de realizar el diseño de forma que la sensibilidad sea máxima en torno al nivel de alarma.

Por una parte, la resistencia del sensor en función de la concentración de hidrógeno, según el ejercicio 9.1, es:

$$R_{[C]} = 1,6 \cdot 10^4 \cdot [C]^{-0,6}$$

y se puede expresar de forma equivalente como:

$$R_{[C]} = R_{100\ ppm} \cdot \left(\frac{[C]}{100\ ppm} \right)^{-0,6}$$

donde $R_{100\ ppm}$ es el valor de la resistencia a la concentración de alarma (100 ppm).

Por otra parte, la tensión de salida del circuito de la figura 9.5:

$$V_o = \frac{R_c}{R_{100\ ppm} \cdot \left(\frac{[C]}{100\ ppm} \right)^{-0,6} + R_c} \cdot V_{cc}$$

y la sensibilidad, S, es:

$$S = \frac{dV_o}{d[C]} = \frac{R_c R_{100\ ppm} \cdot \left(\frac{[C]}{100\ ppm} \right)^{-0,6} \cdot \frac{0,6}{[C]}}{\left[R_{100\ ppm} \cdot \left(\frac{[C]}{100\ ppm} \right)^{-0,6} + R_c \right]^2} \cdot V_{cc}$$

El máximo de esta expresión, siendo [C] = 100 ppm, se puede obtener sin más que derivar la sensibilidad con respecto a la resistencia de carga e igualar a cero, obteniéndose:

$$R_c = R_{100\ ppm}$$

Por lo tanto, se escogerá una resistencia de aproximadamente 1 kΩ. En términos generales, se deberá escoger una resistencia de carga de un valor óhmico igual al de la resistencia del sensor a la concentración de alarma.

Cabe finalmente señalar que existen sensores para la detección de un buen número de gases de interés industrial, entre otros:

- Gases combustibles: propano, metano, hidrógeno, etc.
- Gases tóxicos: monóxido de carbono, amoniaco.
- Clorofluorocarbonos (CFCs).
- Disolventes orgánicos: alcohol, tolueno.

9.2 Magnetorresistencias

La resistencia eléctrica de muchos materiales magnéticos depende del campo magnético aplicado. La razón de tales variaciones de resistencia se debe al espín de los electrones [9.5]: las nubes electrónicas que rodean a los átomos se orientan según el campo magnético y pueden dificultar el paso de la corriente según la posición que adopten (figura 9.6).

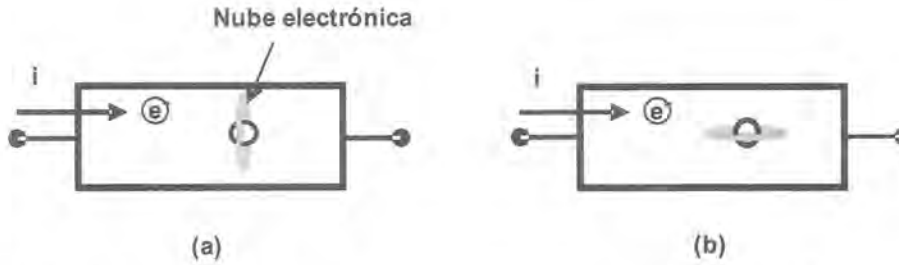


Figura 9.6. Principio físico de las magnetorresistencia: (a) la orientación de la nube electrónica que rodea a los átomos dificulta el paso de la corriente; (b) la resistencia adquiere el mínimo valor puesto que la nube electrónica no dificulta el paso de la corriente.

Las magnetorresistencias vienen magnetizadas en una determinada dirección. Así, en la figura 9.7a se muestra el vector de magnetización, M, y la dirección de la corriente en ausencia de campo externo.

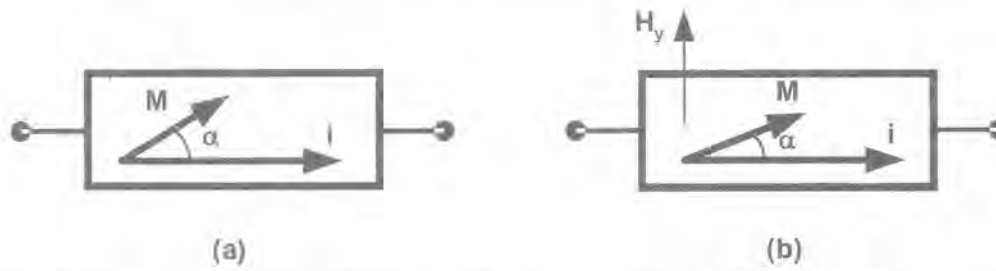


Figura 9.7. (a) Vector de magnetización M en ausencia de un campo magnético externo; (b) rotación del vector de magnetización en presencia de un campo externo.

En presencia de un campo magnético cuya componente en la dirección perpendicular a la dirección de la corriente es H_y (figura 9.7b), el vector de magnetización gira formando un ángulo α con respecto a la dirección de la corriente, dado por [9.6]:

$$\text{sen}^2(\alpha) = \frac{H_y^2}{H_0^2} \tag{9.2}$$

donde H_0 es el campo característico del material debido a la magnetización inicial. La relación entre la resistencia y el campo magnético es:

$$R = R_0 + \Delta R_0 \left(1 - \frac{H_y^2}{H_0^2}\right) \quad (H \leq H_0) \tag{9.3}$$

donde ΔR_0 es la diferencia entre la máxima resistencia posible y la mínima. Según esta última ecuación, la variación de la resistencia con el campo magnético aplicado es parabólica siempre que su orientación sea fija. Si por el contrario se mantiene el campo y se varía su orientación, la resistencia sigue la relación mostrada en la figura 9.8 [9.1]. De todas formas, las magnetorresistencias comerciales presentan geometrías especiales de forma que se trabaje en zonas lineales. Una de las más conocidas es la estructura conocida con el nombre "*barber pole*" [9.6].

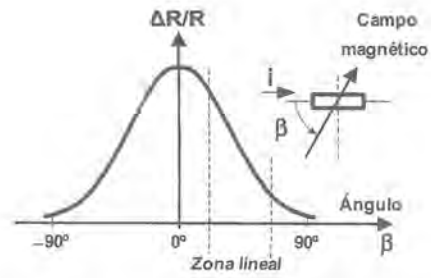


Figura 9.8. Variación de la resistencia con la orientación del campo magnético.

Existen tres tipos de magnetorresistencias:

- Anisotrópicas (*anisotropic magnetoresistor, AMR*): generalmente, están compuestas de una fina película de hierro-níquel (*permalloy*). La variación máxima de la resistencia es del 2-3% en presencia de campo magnético y el ancho de banda llega a 15 MHz.
- Gigantes (*giant magnetoresistor, GMR*): fueron desarrolladas a partir de 1988. Presentan una estructura multicapa obteniéndose variaciones de resistencia de hasta el 50%.
- Colosales (*colossal magnetoresistor, CMR*): de desarrollo aún más reciente, presentan variaciones de resistencia de hasta el 99,9%.

Las magnetorresistencias permiten medir tanto la orientación como la magnitud del campo magnético (desde 0,7 nT hasta 0,6 mT). Algunas aplicaciones son los magnetómetros, detectores de proximidad, etc. Muchas aplicaciones están relacionadas con la medida del cambio del campo magnético terrestre ante la presencia de un material ferromagnético (por ejemplo, detectores de presencia de vehículos) o con la orientación respecto al propio campo magnético terrestre [9.2] (sistemas de navegación, inclinómetros).

Resumen

Los sensores resistivos para la detección de gases presentan una resistencia creciente con la concentración de oxígeno y decreciente, en general, con la concentración del resto de los gases que pueden detectar.

Están compuestos de cuatro terminales: dos corresponden a la resistencia y otros dos a unos electrodos de calentamiento.

Algunos inconvenientes de estos sensores son la escasa selectividad, la dependencia de la temperatura y de la humedad.

Se suelen utilizar para detectar niveles de alarma de gases combustibles.

En las magnetorresistencias, el valor óhmico varía con el valor del campo magnético y con su orientación.

Existen tres tipos de magnetorresistencias: anisotrópicas, gigantes y colosales.

Se utilizan en sistemas relacionados con la medición de campo magnético, como la medida de corriente eléctrica, o con la orientación, como sistemas de navegación e inclinómetros.

Problemas propuestos

9.1 Se dispone de un sensor de hidrógeno cuya resistencia varía con la concentración según la siguiente ecuación (véase el ejercicio 9.1):

$$R_{[C]} = 1,6 \cdot 10^{4} \cdot [C]^{-0,6} \quad \text{a} \quad 20^{\circ}C$$

Se pretende diseñar un sistema de alarma que se active cuando la concentración de hidrógeno sobrepase 100 ppm, por lo que la resistencia de carga será de 1 kΩ (véase el ejercicio 9.2). Suponiendo que el circuito se alimentará a 5 V:

a) Diseñe el circuito que active la alarma.

b) Si el fabricante indica en sus hojas de características que la variación de la resistencia con la temperatura es la indicada en la siguiente gráfica, ¿a qué concentración de hidrógeno se activará la alarma cuando la temperatura sea de 10°C?, ¿y cuando sea de 30°C?

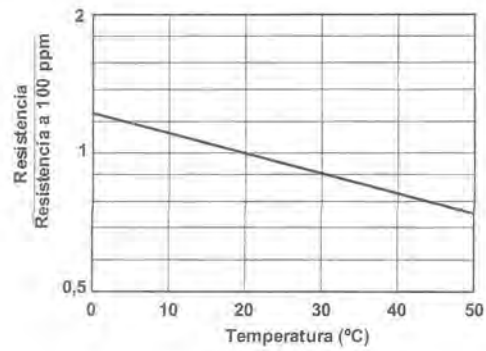


Figura 9.9

9.2 El fabricante de un sensor resistivo para la detección de amoníaco indica en las hojas de características que la resistencia a 50 ppm está comprendida entre 20 kΩ y 100 kΩ. Además, indica que la relación entre la resistencia a 150 ppm y a 50 ppm está comprendida entre 0,4 y 0,7. Proponer una ecuación que refleje la resistencia del sensor en función de la concentración de amoníaco.

9.3 Se pretende realizar un detector de nivel de alarma de un gas mediante un sensor resistivo de óxido de estaño. Proponer un esquema para realizar la compensación de la temperatura utilizando una NTC.

9.4 ¿Cuál es la relación entre la resistencia y la orientación del campo magnético externo en una magnetorresistencia?

10

Sensores capacitivos

10.1 Introducción
10.2 Variación de la capacidad en un condensador de placas paralelas
10.3 Circuitos de medida
10.4 Detectores de proximidad capacitivos
10.5 Sensores capacitivos en silicio

Los sensores capacitivos están basados en la variación de la capacidad entre dos o más conductores entre los que se encuentra un dieléctrico, en respuesta a la variación de alguna magnitud física. Aunque las aplicaciones más conocidas están asociadas con medida de movimiento lineal y angular, los sensores capacitivos se emplean también para la medida de humedad, presión, nivel, aceleración, etc. Las principales características que presentan son:

- Permiten detectar muy pequeños desplazamientos (hasta 10^{-10} mm).
- Son fácilmente integrables en un chip de silicio.
- No se ven afectados por la temperatura o el alineamiento mecánico.
- Son muy estables en entornos hostiles.
- Tienen muy bajo consumo.

Sin embargo, a pesar de estas y otras ventajas que poseen, sus aplicaciones[1] se ven ralentizadas por diversos factores: necesidad de circuitos de acondicionamiento especializados, creencia de que son no lineales y de que no pueden operar con humedades extremas, así como un cierto desconocimiento de la tecnología.

10.1 Introducción

El elemento sensor básico de los sensores capacitivos está constituido por un condensador de capacidad C. El concepto de capacidad fue definido por Maxwell en 1873, en su teoría sobre el campo electromagnético. Si tenemos una configuración con cualquier número de electrodos tal como la mostrada en la figura 10.1, la

[1] Según estudios de mercado realizados en EE.UU. por *Venture Development Corporation* [10.14], las previsiones de las diferentes tecnologías de sensores para el año 2000 fueron: fotoeléctricos (47%), inductivos (42%), magnéticos (6%), capacitivos (3,1%), ultrasonidos (1,9%).

capacidad entre dos de los electrodos (por ejemplo, el i y el j), está dada por el cociente entre la carga inducida en uno de los electrodos y la diferencia de potencial entre los dos electrodos. Es decir:

$$C_{ji} = \frac{Q_{ji}}{V_j - V_i} \tag{10.1}$$

donde C_{ji} es la capacidad entre los electrodos i y j; Q_{ji} es la carga en el electrodo j inducida por la diferencia de potencial $(V_j - V_i)$; V_i es el potencial en el electrodo i; y V_j el potencial en el electrodo j. Esto significa que el potencial de los otros electrodos no contribuye a la capacidad entre los electrodos i y j.

Figura 10.1. Representación fundamental de la capacidad entre conductores.

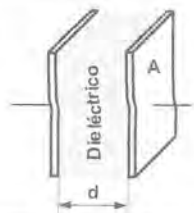
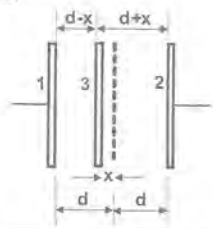
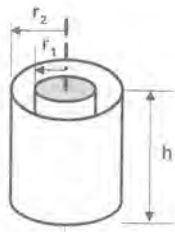
Tipo de condensador	Capacidad
Simple 	$C = \frac{\varepsilon A}{d}$ (10.2) $\varepsilon = \varepsilon_0 \cdot \varepsilon_r$ = permitividad del dieléctrico o constante dieléctrica $\varepsilon_0 = 8{,}854$ pF/m (constante dieléctrica del vacío) ε_r = permitividad relativa (1 para el vacío) A = área de cada placa (m^2) d = distancia entre placas (m)
Diferencial 	$C_1 = \frac{\varepsilon A}{d - x}$; $C_2 = \frac{\varepsilon A}{d + x}$ (10.3) C_1 = Capacidad entre la placa fija 1 y la móvil 3 C_2 = Capacidad entre la placa fija 2 y la móvil 3
Cilíndrico 	$C = \frac{2\pi\varepsilon h}{\ln \frac{r_2}{r_1}}$ (10.4)

Tabla 10.1. Geometrías básicas de condensadores de aplicación en sensores capacitivos.

Para geometrías de electrodos sencillas, el cálculo de la capacidad puede realizarse empleando la Ley de Gauss. Los resultados de la tabla 10.1 se han obtenido de esta forma. Cuando se trata de geometrías más complejas se puede acudir a resolver analíticamente la ecuación de Laplace o bien emplear herramientas software basadas en métodos de elementos finitos. La Universidad de Delf en Holanda tiene un gran número de trabajos sobre sensores capacitivos. En particular, Heerens [10.4] calcula la capacidad de diversas configuraciones de electrodos y su aplicación al diseño de sensores capacitivos.

10.2 Variación de la capacidad en un condensador de placas paralelas

10.2.1 CONDENSADOR SIMPLE

En el caso de un condensador de placas paralelas simple, el valor de la capacidad puede verse alterada variando la distancia entre placas (x), el área de las mismas (A) o el tipo de dieléctrico (ε). Cuando el parámetro que varía es la distancia entre placas se tiene:

$$C = \frac{\varepsilon A}{x} \; ; \; |Z| = \frac{1}{\omega C} = \frac{x}{\omega \varepsilon A} \tag{10.5}$$

Es decir, la capacidad no es lineal con respecto a la distancia entre placas, en cambio, sí lo es la impedancia. La sensibilidad de la capacidad a cambios en la separación de las placas está dada por:

$$\frac{dC}{dx} = -\varepsilon \frac{A}{x^2} \tag{10.6}$$

Vemos cómo la sensibilidad es mayor para pequeños desplazamientos. Para el caso en que lo que varíe sea el área de las placas o el material dieléctrico, la tabla 10.2 muestra los resultados que se obtienen. Vemos que en ambos casos la capacidad varía linealmente con el desplazamiento de la placa, con lo que, desde el punto de vista de la linealidad, se puede medir directamente bien su capacidad o la admitancia correspondiente. Una aplicación típica del sensor capacitivo basado en la variación del área es la medida de desplazamientos angulares, mientras que la variación de la constante dieléctrica puede utilizarse para medir el nivel de un fluido en un tanque.

10.2.2 CONDENSADOR DIFERENCIAL

Se denomina condensador diferencial al formado por tres placas planas paralelas. En general, las placas exteriores suelen ser fijas y la placa central móvil, en respuesta a la variable a detectar (figura 10.2).

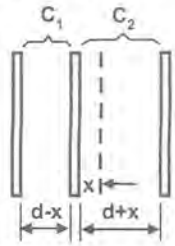


Figura 10.2. Condensador diferencial con variación de la distancia entre placas.

Se forman así dos condensadores de capacidades C_1 y C_2, cuyo valor está dado por:

$$C_1 = \frac{\varepsilon A}{d\text{-}x}\ ;\ C_2 = \frac{\varepsilon A}{d+x} \tag{10.6}$$

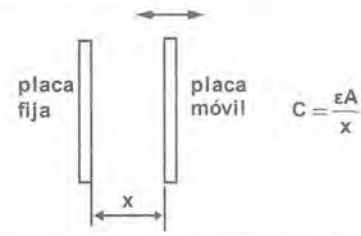
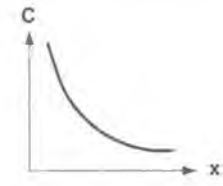
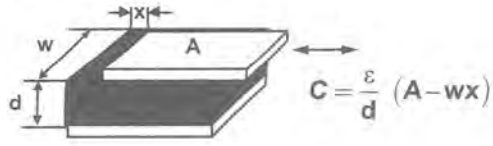
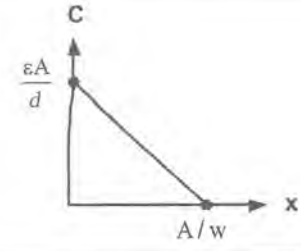
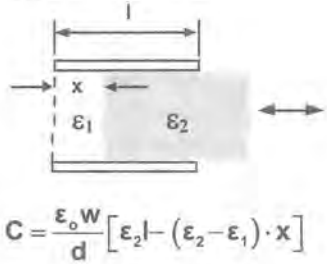
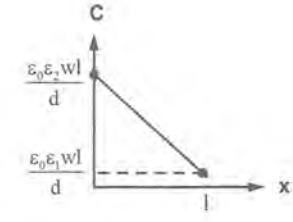
Condensador plano	Función de transferencia
Distancia entre placas variable 	
Área variable 	
Dieléctrico variable 	

Tabla 10.2. Variación de la capacidad en un condensador simple.

La capacidad diferencial está dada por:

$$C_1 - C_2 = \frac{2\varepsilon A}{d^2 - x^2}\cdot x \tag{10.7}$$

Para pequeños valores de desplazamiento x en relación con la distancia entre placas d, la expresión anterior se puede aproximar por:

$$C_1 - C_2 = \frac{2\varepsilon A}{d^2}\cdot x \tag{10.8}$$

Vemos cómo, a pesar de que C_1 y C_2 son no lineales, se logra obtener una salida lineal realizando una medida diferencial y con una sensibilidad mayor que en el caso del condensador simple. Aplicaciones típicas de esta configuración son los micrómetros, acelerómetros e inclinómetros.

Si se disponen las tres placas del condensador diferencial como se muestra en la figura 10.3 y se varía el área entre placas, las capacidades C_1 (capacidad entre las placas 1 y 3) y C_2 (capacidad entre las placas 2 y 3) están dadas por:

$$C_1 = \frac{\varepsilon}{d} w(x_0 - x);\ C_2 = \frac{\varepsilon}{d} w(x_0 + x) \tag{10.9}$$

donde w es el ancho de las placas, definido anteriormente en la tabla 10.2. La capacidad diferencial $C_1 - C_2$ está dada por:

$$C_1 - C_2 = -\frac{2\varepsilon}{d} \cdot wx \tag{10.10}$$

De nuevo la medida diferencial de la capacidad nos permite obtener una salida lineal con el desplazamiento. Con esta técnica se han llegado a medir desplazamientos de hasta 10^{-10} mm [10.5].

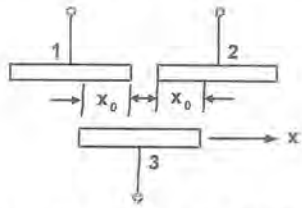


Figura 10.3. Variación del área entre placas en un condensador diferencial.

Aunque se ha mejorado en linealidad, el condensador diferencial no está exento de problemas: dispersión del campo eléctrico en los bordes de los electrodos, aislamiento entre placas, capacidades parásitas, cables de conexión, etc. La dispersión del campo en los bordes de los electrodos puede apreciarse mejor en la figura 10.4a. Para evitar este efecto, en la figura 10.4b se ha colocado un **electrodo de guarda** rodeando al electrodo 3. Este electrodo de guarda se mueve junto al electrodo 3 y está puesto a su mismo potencial. Ahora, las líneas de campo aparecen distorsionadas en el borde de la guarda, lo cual no importa ya que no es un electrodo de medida. En cambio el campo sí es uniforme en los electrodos de medida. El electrodo de guarda protege simultáneamente al electrodo 3 de campos externos que puedan afectar a la medida.

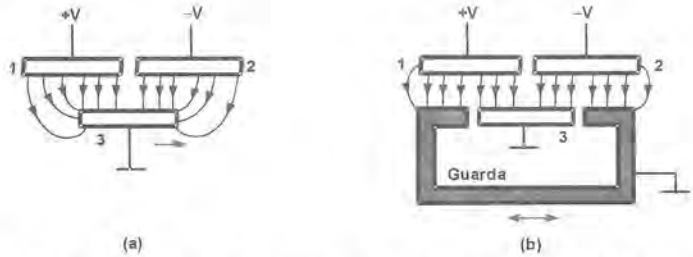


Figura 10.4. a) Dispersión del campo eléctrico en los bordes de los electrodos; b) electrodo de guarda para reducir el efecto de los bordes.

10.3 Circuitos de medida

Una vez calculada la capacidad entre los electrodos del sensor, llega el momento de ver cómo medir dicha capacidad. Se trata de convertir las variaciones de capacidad, debidas a la variable a medir, en una tensión, frecuencia o modulación de ancho de pulso. En primer lugar empezaremos seleccionando la señal de excita-

ción del sensor capacitivo y, a continuación, haremos un breve recorrido por los diversos circuitos y tecnologías que se han ido proponiendo para su acondicionamiento.

10.3.1 SEÑAL DE EXCITACIÓN

Al tratarse de una impedancia, la señal de excitación de un sensor capacitivo ha de ser obligatoriamente una señal alterna. El valor de la frecuencia de excitación ha de ser tal que la impedancia no tenga un valor excesivo. Por ejemplo, en un condensador de placas planas con A = 1 cm^2 y d = 1 mm, la capacidad que resulta aplicando la expresión (10.2) es C = 0,8 pF y el valor de la impedancia para una frecuencia de excitación de 100 kHz es aproximadamente 2 MΩ. Los valores de capacidad típicos suelen estar en el margen de 0,1 a 500 pF y la frecuencia de excitación nunca suele ser inferior a 100 kHz. Si se eleva la frecuencia excesivamente se tienen dos problemas: por un lado, el consumo de los circuitos se incrementa y, por otro lado, la demodulación de la señal se hace más difícil.

En cuanto a la forma de onda se refiere, en circuitos que requieran tener una alta precisión se prefieren formas de onda senoidales, especialmente cuando los valores de capacidad son extremadamente pequeños que obliga a elevar la frecuencia por encima del megahercio. La ventaja que tiene con respecto a formas de onda cuadradas es que las limitaciones del *slew rate* del amplificador y las no linealidades, consecuencia del desplazamiento de fase, son problemas menos importantes. Además, las ondas senoidales producen menos interferencias electromagnéticas (EMI), que las formas de ondas cuadradas. Si bien las ondas senoidales no requieren tener unas características de distorsión excepcionalmente buenas, sí se requiere que la estabilidad de la amplitud sea alta, por lo que la generación de dichas formas de ondas no es una tarea fácil. En estas condiciones, los osciladores senoidales clásicos como el oscilador de Colpitts y el de puente de Wien no son adecuados. Una posible solución es generar una onda cuadrada a partir de un cristal de cuarzo y después realizar un filtrado paso bajo, analógico o digital.

10.3.2 AMPLIFICADOR DE ALTERNA

La solución más directa es emplear como circuito básico de medida de la capacidad el amplificador de alterna de la figura 10.5. La tensión de salida es proporcional a C_1/C_2:

$$V_o = -V \cdot \frac{C_1}{C_2} \tag{10.11}$$

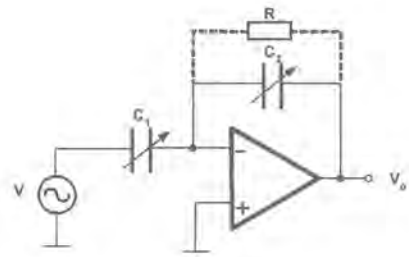


Figura 10.5. Amplificador de alterna para la medida de la capacidad.

Este acondicionamiento puede utilizarse para obtener una salida lineal con la variación del área de las placas o con la distancia de separación entre ellas. En el primer caso C_1 es la capacidad que experimenta la variación del área de las placas y C_2 una capacidad de referencia. En el segundo caso C_1 es el condensador

de referencia y C_2 el que varía la distancia entre placas. La resistencia paralelo R permite la circulación de las corrientes de polarización del operacional (véanse capítulos 2 y 3). En la práctica este circuito presenta problemas de estabilidad, por lo que no es una solución adecuada.

10.3.3 PUENTES DE ALTERNA

Los cambios en la capacidad de los sensores capacitivos pueden ser medidos empleando puentes de alterna. Estos puentes son, como veremos, especialmente atractivos en el caso del condensador diferencial. La figura 10.6 muestra el caso de un puente de alterna en el que se disponen los condensadores C_1 y C_2 en ramas adyacentes, estando ocupadas las otras dos ramas por resistencias de igual valor.

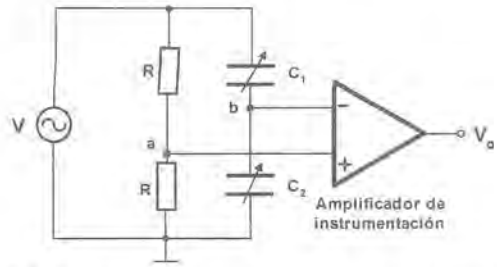


Figura 10.6. Puente de alterna para la medida de la capacidad.

La tensión en extremos del puente es:

$$V_{ab} = V\left(\frac{1}{2} - \frac{C_1}{C_1 + C_2}\right) \tag{10.12}$$

Sustituyendo las expresiones de C_1 y C_2 dadas por (10.9), resulta:

$$V_{ab} = V\frac{x}{2x_0} \tag{10.13}$$

Si la salida del puente se detecta con un circuito de alta impedancia de entrada (por ejemplo, mediante un amplificador de instrumentación), la relación entre la salida del puente y el desplazamiento es lineal. La estructura en puente aporta una ventaja adicional, ya que cualquier cambio que se produzca de forma simultánea en las ramas adyacentes se cancela (por ejemplo, interferencias externas, derivas térmicas, etc.). Por el contrario, los brazos resistivos del puente introducen errores debidos a las capacidades parásitas de dichas resistencias. Sus efectos se pueden reducir mediante un puente cuyos brazos sean inductivos, como se muestra en la figura 10.7.

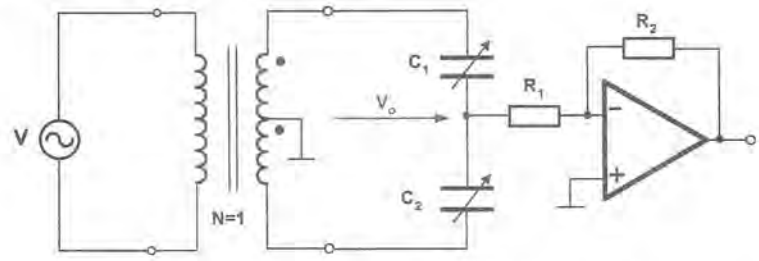


Figura 10.7. Puente capacitivo con transformador.

La tensión de salida del puente V_o, está dada por:

$$V_o = \frac{V}{2} \cdot \frac{C_1 - C_2}{C_1 + C_2} \tag{10.14}$$

Si varía la distancia entre placas en el sensor capacitivo diferencial, llevando a la expresión (10.14) los valores de C_1 y C_2 dados por (10.9), se comprueba que la salida es lineal con el desplazamiento, de la forma:

$$V_o = \frac{V}{\varepsilon x_0} \cdot x \tag{10.15}$$

10.3.4 OSCILADOR DE FRECUENCIA VARIABLE

En muchas ocasiones el sensor capacitivo forma parte de un circuito oscilador, en el que las variaciones en la capacidad provocan una variación en la frecuencia de oscilación. Existen diferentes soluciones para implementar el circuito oscilador. La figura 10.8 muestra un oscilador RC, con ciclo de trabajo del 50%, a base del venerable 555 en su versión CMOS de bajo consumo [10.5]. La frecuencia de oscilación de la salida del 555 es: $f_{OSC} = 1/1,4RC$ [10.7].

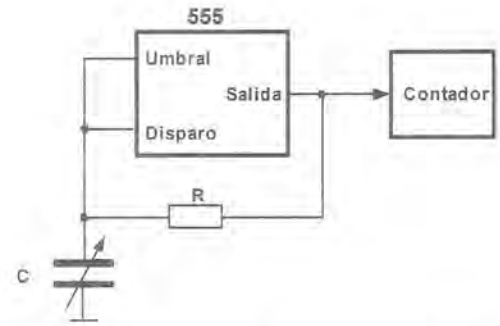


Figura 10.8. Oscilador RC mediante el CI 555.

Si el condensador varía su distancia entre placas, la frecuencia de oscilación del circuito será proporcional a dicha distancia y la salida del sensor será lineal si se cuenta el número de pulsos del oscilador. Si lo que varía es el área de las placas, la salida se linealizará midiendo la duración de los pulsos.

Este circuito no está exento de problemas. Por un lado, la capacidad parásita del cable de conexión del condensador se añade a la capacidad de medida, reduciendo la sensibilidad. Otro inconveniente se debe a la resistencia R de valor fijo que interviene en la medida de la capacidad. Su estabilidad y coeficiente de temperatura pueden no ser lo suficiente buenos, afectando también a la medida. Estos problemas se pueden reducir empleando el circuito demodulador síncrono que se presenta a continuación.

10.3.5 DEMODULADOR SÍNCRONO

Este método es el que permite medir la capacidad con mayor flexibilidad y precisión. La figura 10.9 muestra su esquema básico. El demodulador síncrono se comporta en esencia como un rectificador de doble onda, realizando dicha rectificación en sincronismo con la señal de excitación o de sincronismo del circuito. Así, durante el primer semiciclo, el valor de la señal de excitación es superior a una referencia y el demodulador

funciona como un rectificador inversor. Durante el segundo semiciclo, la señal de excitación es inferior a la referencia y el demodulador funciona como un seguidor de tensión, de forma que la entrada al demodulador aparece en la salida con ganancia unidad. En la figura 10.9 se muestra con línea de puntos y continua lo que ocurre en el demodulador, dependiendo del sentido de desplazamiento que se tenga en el condensador.

El filtro paso bajo elimina la frecuencia de la portadora y otros armónicos de orden superior produciendo una salida libre de rizado. La amplitud de la señal rectificada y filtrada informa de la magnitud del desplazamiento, mientras que la polaridad informa del sentido del desplazamiento. Dependiendo de la configuración del amplificador se puede linealizar bien la capacidad o la impedancia.

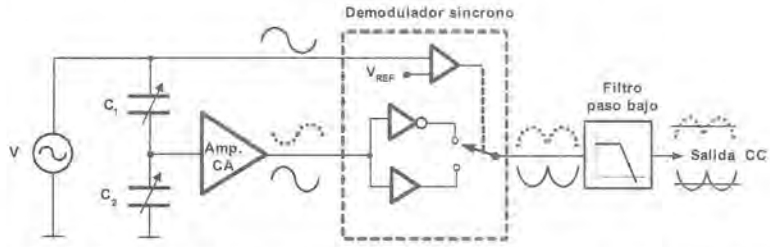


Figura 10.9. Principio de funcionamiento de un circuito demodulador síncrono.

Existen circuitos comerciales como el NE5521 de Philips o el AD698 de Analog Devices que integran oscilador, amplificador, demodulador y filtro. Si bien estos circuitos fueron pensados inicialmente para el acondicionamiento del LVDT (véase capítulo 11), pueden utilizarse también con un sensor capacitivo de placas paralelas simple o diferencial. Así, la figura 10.10 muestra la conexión de un condensador diferencial al NE5521 [10.10]. La frecuencia de oscilación está dada por: $f_{OSC} = (V_{REF} - 1{,}3\ V)/[V_{REF}(R_T + 1{,}5\ k)\cdot C_T]$. El amplificador auxiliar del NE5521 permite implementar el filtro paso bajo de la figura 10.9.

10.3.6 CIRCUITOS DE CAPACIDAD CONMUTADA

La tecnología de capacidad conmutada está presente en los sensores capacitivos en silicio. El resultado es una reducción muy importante del área de silicio necesaria para integrar resistencias de valor elevado, en comparación con las técnicas de difusión. Para ello se sustituyen dichas resistencias por interruptores CMOS y condensadores, en los que la disipación de potencia es insignificante.

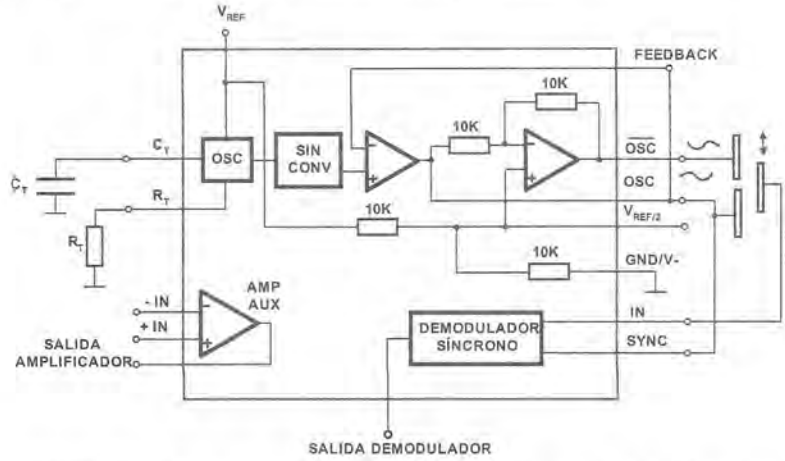


Figura 10.10. Configuración del circuito NE5521 como acondicionador de señal en un sensor capacitivo.

En la figura 10.11a se muestra el esquema de un integrador convencional y en la figura 10.11b el mismo integrador, pero realizado con tecnología de capacidad conmutada, donde la resistencia R se ha sustituido por dos interruptores MOS (S_1 y S_2) y un condensador C. El funcionamiento de este circuito se explicó con detalle en el capítulo 4. Se obtuvo para "R" un valor dado por:

$$"R" = \frac{1}{C_1 f_{CLK}} \tag{10.16}$$

Es decir, el valor de "R" puede controlarse ajustando durante el proceso de fabricación los valores de la capacidad C_1 y de la frecuencia del reloj, obteniéndose así valores de resistencias equivalentes con mucha precisión.

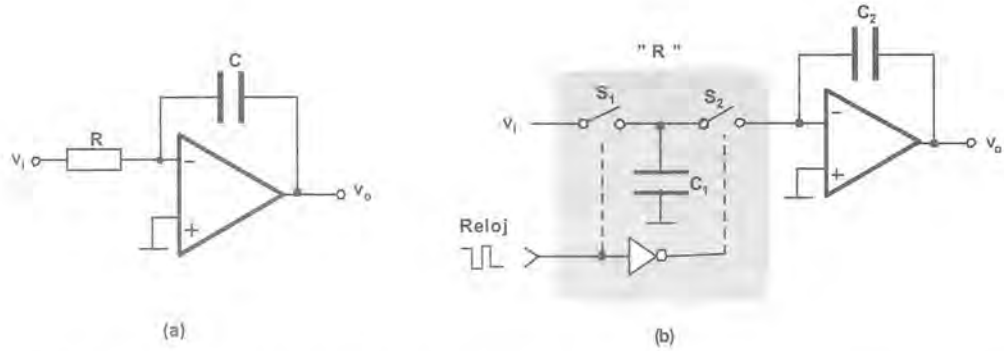


Figura 10.11. a) Integrador convencional; b) integrador de capacidad conmutada.

Mediante técnicas de capacidad conmutada pueden implementarse los circuitos de acondicionamiento que hemos estado considerando. La figura 10.12 muestra un amplificador diferencial de capacidad conmutada [10.1]. Un reloj de dos fases ϕ_1 y ϕ_2, abre y cierra de forma alternativa los interruptores a una frecuencia lo suficientemente grande para que tanto V_1 como V_2 no cambien de forma apreciable en cada periodo de conmutación. Cuando los interruptores gobernados por la fase ϕ_1 se cierran, C_1 se carga con $V_1 - V_2$ voltios, almacenando la carga $q_1 = C_1(V_1 - V_2)$. Cuando se cierran los interruptores gobernados por la fase ϕ_2, la masa virtual hace que C_1 se descargue a cero voltios, transfiriendo a C_2 toda su carga almacenada. La tensión de salida está dada por:

$$V_0 = \frac{C_1}{C_2} \cdot (V_1 - V_2) \tag{10.17}$$

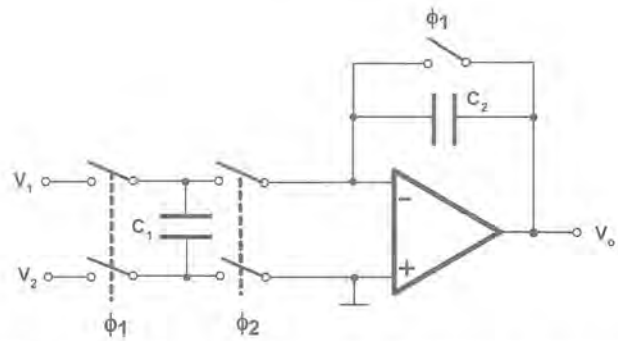


Figura 10.12. Amplificador diferencial de capacidad conmutada.

La figura 10.13 muestra la implementación de un amplificador de instrumentación mediante el integrado LTC1043 de Linear Technology [10.5]. Este circuito está compuesto por ocho interruptores, tiene una razón de rechazo de modo común de 120 dB y una frecuencia máxima de reloj de 5 MHz. La ganancia del circuito es $1 + R_2/R_1$. El condensador C_T es el que establece la frecuencia de conmutación. La ventaja de esta solución con respecto al amplificador de instrumentación tradicional es una mayor razón de rechazo de modo común.

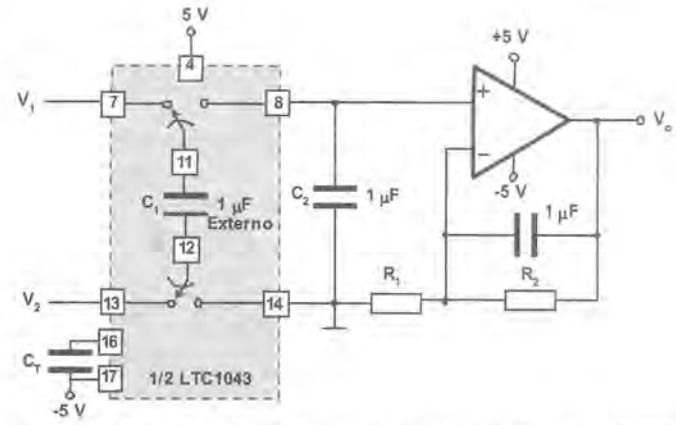


Figura 10.13. Amplificador de instrumentación a partir del LTC1043.

10.4 Detectores de proximidad capacitivos

Los detectores de proximidad constituyen una de las principales aplicaciones de los sensores capacitivos. Detectan la presencia de objetos próximos, tanto metálicos como no metálicos sin requerir contacto directo con dichos objetos. Se encuentran presentes en una gran variedad de aplicaciones industriales: control de nivel de líquidos, inspección de paquetes, detección de rotura de cables, detección de nivel en productos de alimentación, detección de pequeños elementos metálicos, etc.

La figura 10.14 muestra el aspecto constructivo de un detector de proximidad capacitivo, así como el diagrama de bloques interno. Comercialmente se encuentran en forma cilíndrica (lisa o roscada), con el cuerpo metálico o plástico y en forma rectangular con la cabeza sensora orientable. La cabeza sensora contiene como mínimo, un par de electrodos que constituyen las placas de un condensador abierto. Estos electrodos están situados en el lazo de realimentación de un oscilador de alta frecuencia. Cuando no se tiene presente un objeto, la capacidad del sensor es baja y, por tanto, la amplitud de la oscilación es pequeña. Cuando el objeto se aproxima a la cabeza sensora se incrementa la capacidad y, con ello, la amplitud de la oscilación, provocando la conmutación del circuito de salida del detector.

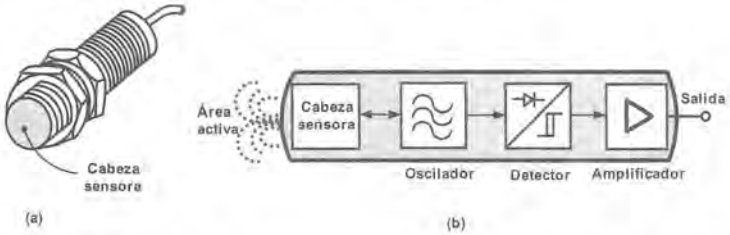


Figura 10.14. Detector de proximidad capacitivo: (a) aspecto constructivo; (b) diagrama de bloques interno.

Si el objeto es conductor (figura 10.15a) se formará un condensador entre el objeto y la cara activa del sensor, de forma que cuando el objeto se acerca se incrementa la capacidad. Si el objeto no es metálico (sólido o líquido), actúa como un aislante entre los electrodos A y B. En este caso, al aproximarse al detector, lo que ocurre es un incremento de la permitividad dieléctrica del medio y, por tanto, de la capacidad. Recordemos que todos los líquidos y sólidos no conductores tienen una permitividad dieléctrica relativa mayor que la del aire. En el apéndice 10.1 al final de este capítulo se dan las permitividades dieléctricas de algunos productos industriales. En el caso de materiales conductores (metales, líquidos, etc.) se debe conectar el objeto a detectar a tierra a fin de anular la capacidad parásita entre el objeto y tierra.

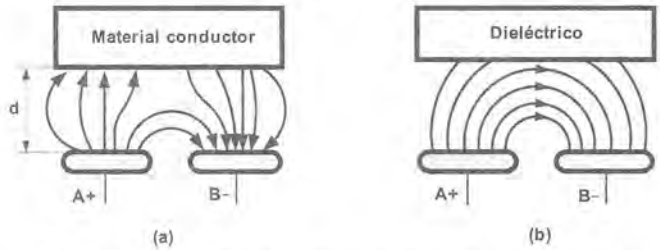
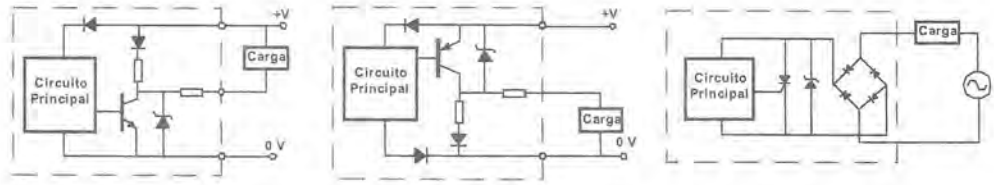


Figura 10.15. Influencia del tipo de material sobre el campo eléctrico.

Como ejemplo de interruptor de proximidad capacitivo, la tabla 10.3 muestra las características más significativas de los interruptores de proximidad de la familia E2K-X de OMRON [10.8].

Características técnicas						
Tipos	c.c, tres hilos			c.a, dos hilos		
Tamaños	M12	M18	M30	M12	M18	M30
Alcance nominal	4 mm	8 mm	15 mm	4 mm	8 mm	15 mm
Objeto	Conductores y dieléctricos					
Tensión de alimentación	10 - 40 Vc.c.			90 - 250Vc.a.		
Respuesta en frecuencia	100 Hz mínimo			10 Hz mínimo		
Consumo	8 mA a 12 V			2,2 mA a 250 Vac, 50 Hz		
Capacidad de interrupción	200 mA máx.			De 10 a 200 mA		
Tipo de salida	Interruptor de c.c. (NPN o PNP)			Interruptor de c.a.		
Diagrama del circuito de salida:						



Ejemplo: Detección del nivel de agua

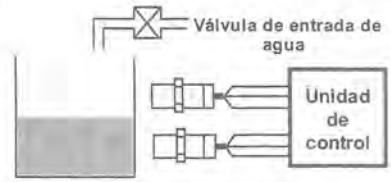


Tabla 10.3. Características del detector de proximidad capacitivo E2K-X de OMRON.

El **alcance nominal (Sn)** que especifica el fabricante no tiene en cuenta las dispersiones de fabricación, temperatura y tensión. Para su determinación se utiliza una placa estándar de acero dulce, con forma cuadrada, de lado 3Sn y un milímetro de espesor.

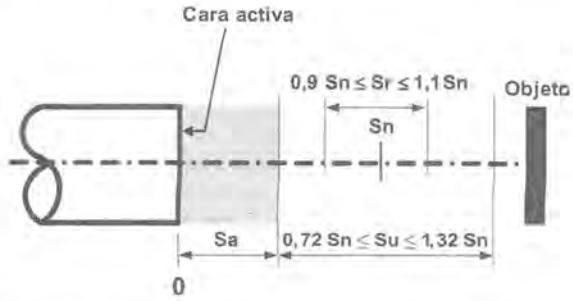


Figura 10.16. Definición de alcances en un detector de proximidad.

Además del alcance nominal, sobre la figura 10.16 se definen los siguientes alcances:

- **Alcance efectivo (Sr):** es el alcance medido en unas condiciones de tensión y temperatura determinadas.
- **Alcance útil (Su):** es el alcance medido en condiciones de temperatura ambiente y tensión de alimentación. Su valor debe estar entre el 80 y 120% del alcance efectivo.
- **Alcance de trabajo (Sa):** es el campo de funcionamiento del aparato. Corresponde al espacio en el cual la detección de la placa de medida es segura, sin importar las dispersiones de tensión y de temperatura. Su valor está comprendido entre 0÷72% del alcance nominal.

10.5 Sensores capacitivos en silicio

La tecnología de fabricación de circuitos integrados en silicio está siendo empleada de forma masiva en el desarrollo de sensores (microsensores de silicio) [10.4]. La integración en silicio permite sustanciales ventajas: mayor sensibilidad, reducción de peso, ahorro en los costes de fabricación, mayor número de aplicaciones, etc. Otras ventajas adicionales son: reducción de los problemas derivados de la temperatura, humedad, capacidades parásitas y de los terminales, además de proporcionar alta impedancia de entrada. El papel que juega el silicio no siempre es el mismo: en unos casos se aprovecharán sus excelentes propiedades elásticas y de estabilidad para detectar una determinada magnitud física, mientras que, en otros casos, el silicio únicamente servirá como un sustrato, siendo necesario añadir finas capas de otros materiales para producir el efecto deseado. Veamos algunos ejemplos significativos de sensores capacitivos en silicio.

Sensor de humedad

Los sensores capacitivos de humedad relativa[2] (RH) se usan en numerosas aplicaciones industriales. Es necesario controlar la humedad en la industria textil, maderera, alimentación, fabricación de papel, almacenaje, etc.

[2] La humedad relativa es la relación entre la presión parcial del vapor de agua presente y la presión del vapor de agua necesaria para alcanzar la saturación a una cierta temperatura. Se expresa en tanto por ciento.

Existen muchos tipos y variantes de sensores de humedad, según sea la composición de las placas del condensador, el material del dieléctrico y el sustrato. El dieléctrico puede ser un óxido de aluminio, silicio poroso o un material polímero. El sustrato es normalmente cerámico, de cristal o de silicio. La figura 10.17 muestra la estructura de un sensor de humedad en el que el dieléctrico es un material polímero y el sustrato es de silicio [10.11]. El material polimero tiene la habilidad de absorber moléculas de agua lo cual se traduce en un cambio en la permitividad dieléctrica del condensador. En los sensores con dieléctrico de silicio la porosidad del silicio puede ser controlada en un proceso de anodizado.

El cambio incremental en la permitividad dieléctrica de un sensor de humedad capacitivo es proporcional a la humedad relativa del entorno que le rodea. La variación de la capacidad se encuentra entre 0,2 y 0,5 pF por 1%RH. Los sensores capacitivos de humedad se caracterizan por un coeficiente de temperatura bajo, capacidad para trabajar a altas temperaturas (hasta 200°C), recuperación rápida de condensación y una buena compatibilidad química. Una característica importante es su estabilidad química ya que en algunas aplicaciones, el vapor de agua contiene contaminantes (por ejemplo, monóxido de carbono) o bien la medida de humedad se realiza en una muestra de gas en lugar de aire. En estos casos, si el dieléctrico es silicio, las características del sensor no se ven afectadas por dichos gases.

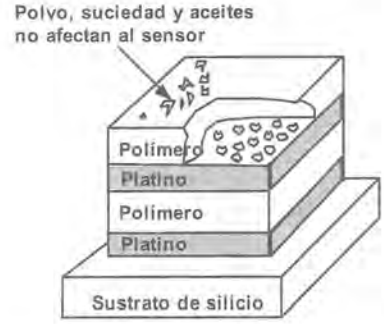


Figura 10.17. Construcción de un sensor de humedad capacitivo.

La circuitería de acondicionamiento consiste normalmente en un oscilador CMOS que excita el sensor, obteniendo una relación lineal entre la tensión de salida y la humedad relativa. Como ejemplo de sensor de humedad capacitivo, la tabla 10.4 muestra las principales características del sensor de humedad HIH-3610 de Honeywell [10.11].

Característica	
Alimentación	5 V c.c.
Margen de medición	0 - 100%RH
Salida	de 0,8 a 3,9 V d.c. (V_{cc} = 5 V, 25°C)
Exactitud	±2%RH
Linealidad	±0,5%RH
Histéresis	±1,2%RH
Repetibilidad	±0,5%RH
Tiempo de respuesta	15 s con aire en calma a 25°C
Estabilidad	±1%RH al 50%RH durante 5 años
Tensión de salida	$V_o = V_{cc}[0,0062 \times (\%HR)+0,16]$ @ 25°C
Compensación de temperatura	%HR real = (%HR)/(1,0546–0,00216T), T en °C

Tabla 10.4. Características del sensor HIH-3610 de Honeywell.

Sensor de presión

Un sensor de presión capacitivo responde, en general, a una estructura como la mostrada en la figura 10.18, en la que se tiene una placa metálica fija y un diafragma flexible. Cuando el diafragma se encuentra sometido a una presión (P), el desplazamiento vertical (y), en un punto cualquiera de radio (r) está dado por:

$$y=\frac{3\left(1-\nu^2\right)\left(a^2-r^2\right)}{16Et^3}\cdot P \tag{10.18}$$

donde: **a** es el radio del diafragma, **t** el espesor del diafragma, **E** el módulo de Young y **ν** la relación de Poisson.

La deformación del diafragma da lugar a que la separación media de las placas se reduzca, lo que se traduce en un incremento relativo de la capacidad dado por la expresión [10.8]:

$$\frac{\Delta C}{C_0}=\frac{\left(1-\nu^2\right)\cdot a^4}{16Edt^3}\cdot P \tag{10.19}$$

siendo d la separación de las placas y $C_0 = \varepsilon_0 \pi a^2/d$, la capacidad a presión cero.

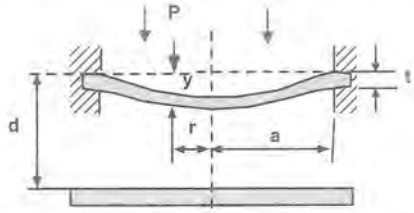


Figura 10.18. Sensor de presión capacitivo de dos placas.

La estructura de la figura 10.18 puede integrarse en silicio, como se muestra en la figura 10.19 [10.11]. El elemento sensor capacitivo está compuesto por un diafragma cerámico, de forma que cuando la presión aumenta, el diafragma se flexiona, variando la distancia entre las placas del condensador. Cambiando el espesor del diafragma en el proceso de fabricación, se pueden llegar a medir presiones entre 7,5 y 10000 psi.

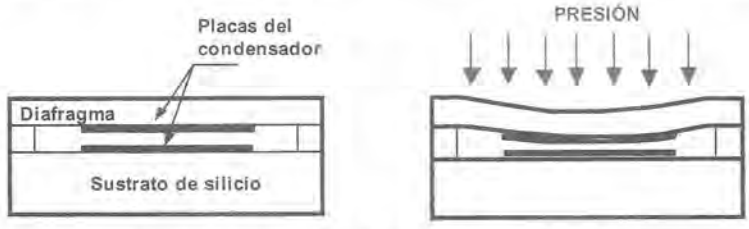


Figura 10.19. Sensor de presión capacitivo.

El circuito integrado incorpora la circuitería de acondicionamiento necesario para convertir la variación de capacidad en una tensión linealmente proporcional a la presión. En muchos casos, la salida de tensión presenta una característica ratiométrica, como se muestra en la figura 10.20a. Esta característica se utiliza en la conexión del sensor a un convertidor A/D. La figura 10.20b muestra dicha conexión, en la que la tensión de alimentación se utiliza como tensión de referencia.

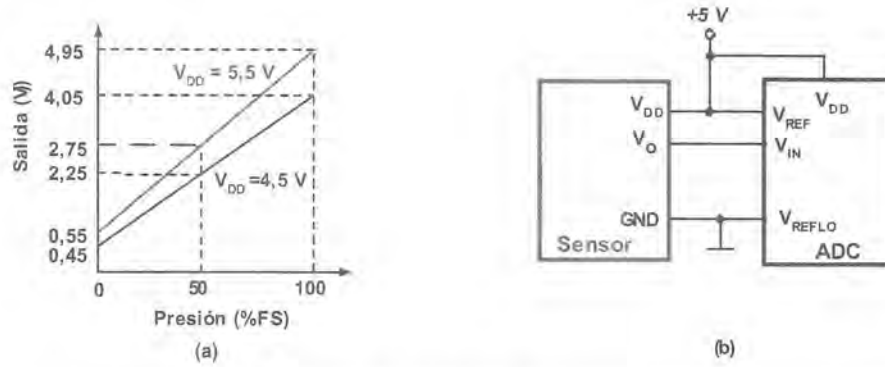


Figura 10.20. a) Característica ratiométrica de un sensor de presión; b) conexión a un convertidor A/D.

Acelerómetro capacitivo

Los acelerómetros se usan para medir aceleración, inclinación, fuerzas inerciales, choques y vibraciones en aplicaciones de automoción (*airbag*), medicina, control industrial, etc. Las últimas técnicas de micromecanizado superficial junto con tecnologías CMOS de fabricación de circuitos integrados permiten integrar en el mismo chip toda la circuitería de procesamiento de la señal, resultando un dispositivo de altas prestaciones y de coste razonable.

La medida de aceleración está basada en la Ley de *Hook*: F = kx y en la segunda Ley de Newton: F = ma. Igualando ambas expresiones, tenemos que la aceleración está dada por la expresión:

$$a = \frac{k}{m} x \tag{10.20}$$

con lo que la medida de la aceleración se convierte en la medida de un desplazamiento.

Un ejemplo de microacelerómetro capacitivo es el ADXL150 de Analog Devices [10.2]. La figura 10.21 muestra una vista simplificada del mismo. Se trata de una masa inercial en forma de H sensible a la aceleración. Consta de 42 celdas unitarias, como la mostrada en la figura, en la que cada celda está compuesta por dos placas fijas y una placa móvil que se deforma en respuesta a una aceleración. Se tiene un condensador diferencial, cuya variación de capacidad se mide en el propio chip. La separación de las placas es de 2 μm, así como el espesor de las mismas, resultando un elemento sensor de 1 mm^2.

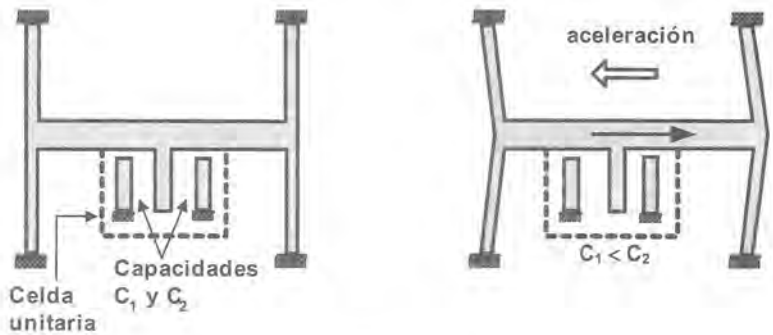


Figura 10.21. Respuesta del acelerómetro a una aceleración aplicada.

El ADXL150 mide la diferencia entre los dos condensadores, la cual está dada por (10.8):

$$\Delta C = \frac{2\varepsilon A}{d^2} \cdot x \tag{10.21}$$

Por tanto, la diferencia en capacidad es proporcional a x, pero sólo para pequeños valores de desplazamiento en comparación con la distancia entre placas.

Toda la circuitería necesaria para el control del sensor y para convertir el cambio de capacidad a tensión está incorporada en el chip, no necesitando componentes externos. La tensión de salida es función de la aceleración a, y de la tensión de alimentación V_S:

$$V_o = \frac{V_S}{2} - (\text{Sensibilidad} - \frac{V_S}{5V} \cdot a) \tag{10.22}$$

La figura 10.22 muestra el diagrama de bloques del acelerómetro ADXL150: la realimentación negativa asegura que el movimiento de las masas es de un valor muy pequeño como para que se pueda considerar correcta la expresión (10.22). El circuito incluye también un filtro paso bajo de Bessel de capacidad conmutada de 2 polos, con un frecuencia de corte de 1 kHz. Los filtros de Bessel tienen una sobreoscilación muy pequeña y un retardo de grupo constante en la banda de paso. La patilla de test se activa con un nivel alto y permite comprobar el funcionamiento del acelerómetro, al proporcionar en la salida una tensión correspondiente al 20% de la aceleración de fondo de escala. El margen de medida está entre ±5 y ±50g, donde g es la aceleración de la gravedad. El error de linealidad es del 0,2% del fondo de escala. La sensibilidad es también ratiométrica y su valor es de 33 mV/g para una V_S = 5 V. El ADXL250 tiene dos ejes sensibles.

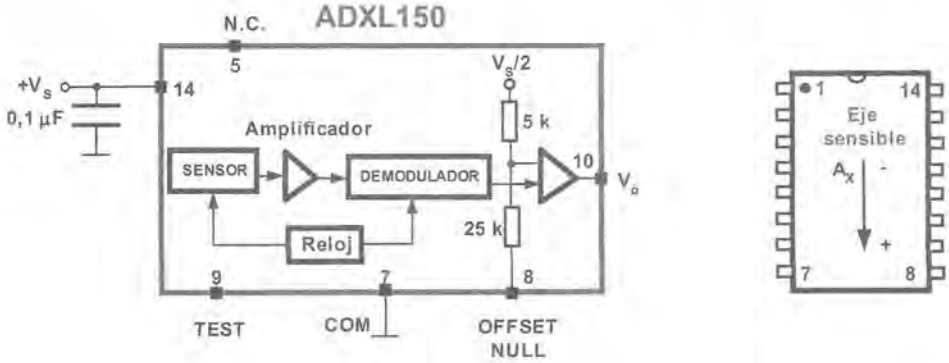


Figura 10.22. Diagrama de bloques funcional del acelerómetro ADXL150.

Resumen

- Los sensores capacitivos basan su funcionamiento en la variación de la capacidad entre dos o más conductores entre los que se encuentra un dieléctrico, en respuesta a la variación de alguna magnitud física.
- Se caracterizan principalmente por ser capaces de detectar pequeños desplazamientos (hasta 10^{-10} mm), fácilmente integrables en un chip de silicio, no verse afectados por la temperatura o el alineamiento mecánico, buena estabilidad en entornos hostiles, así como muy bajo consumo. A pesar de estas buenas características, su cuota de mercado es muy reducida en comparación con otros sensores, como los inductivos o los fotoeléctricos.

- Las geometrías de condensadores más habituales son las de placas planas paralelas configuradas como un condensador simple (dos placas) o diferencial (3 placas). El condensador diferencial tiene una mayor sensibilidad que el condensador simple y permite obtener una salida lineal.
- Al tratarse de una impedancia, la señal de excitación de un sensor capacitivo ha de ser obligatoriamente alterna. Los valores de capacidad suelen estar en el margen de 1 a 500 pF y la frecuencia de excitación nunca suele ser inferior a 100 kHz, de forma que la impedancia sea manejable.
- Entre las diversas técnicas y circuitos empleados para la medida de la capacidad, la demodulación síncrona es la más flexible de todas y con la que se obtiene una mayor precisión, reduciendo los problemas derivados de las capacidades parásitas y del ruido. Esto se ve favorecido por la existencia de circuitos integrados que incorporan todos los elementos necesarios para realizar dicha demodulación síncrona.
- Las principales aplicaciones de los sensores capacitivos en forma discreta son como detectores de proximidad. En forma de circuito integrado se emplean para la medida de diversas variables físicas, principalmente desplazamiento, humedad, presión, nivel y aceleración. Su integración en silicio permite reducir los problemas derivados de la temperatura, humedad, capacidades parásitas, terminales, alta impedancia de entrada, etc.

APÉNDICE 10.1
PERMITIVIDAD RELATIVA DE ALGUNOS PRODUCTOS INDUSTRIALES [10.12]

Material	ε_r	Material	ε_r
Aceite de transformador	2,2 - 2,4	Leche en polvo	3,5 - 4
Acetona	19,5	Madera húmeda	10 - 30
Ácido acético	4,1	Madera seca	2 - 7
Ácido sulfúrico	84	Maíz	5 - 10
Agua	48 - 88	Mica blanca	4,5 - 9,6
Aire	1,0	Nylon	4 - 5
Alcohol industrial	25,8	Neopreno	6 - 9
Amoniaco	15 - 25	Nitrobenceno	36
Anilina	6,9	Óxido de zinc	1,7
Arena	3 - 5	Óxido ferroso	14,2
Arroz	3,5	Papel	1,6 - 2,6
Azúcar	3,0	Parafina	1,9 - 2,5
Baquelita	3,6	Peróxido de hidrógeno	84,2
Benceno	2,3	Petróleo	2,0 - 2,2
Butano	1,4	Poliamida	2,5
Carbón	1,2 - 1,8	Polipropileno	2,0 - 2,3
Caucho	2,5 - 35	Poliéster	2,8 - 8,1
Celulosa	3,2 - 7,5	Policarbonato	2,9
Cemento	1,5 - 2,1	Polivinilo	2,8 - 3,1
Cera	2,4 - 6,5	Porcelana	4,4 - 7
Cereales	3 - 5	Propano líquido	1,6
Vidrio	3,1 - 10	Tolueno	2,3

Material	ε_r	Material	ε_r
Dióxido de carbono	1,6	Vaselina	2,2 - 2,9
Dióxido de silicio	4,5		
Ebonita	2,7 - 2,9		
Etanol	24		
Flúor	2,5 - 3,0		
Freón líquido	6,1		
Gasolina	2,2		
Glicerina	47		
Hielo, –5°C	2,85		
Hielo, –18°C	3,16		

Problemas propuestos

10.1 Se desea diseñar un condensador de placas paralelas cuya capacidad sea de 1 pF y cuyo dieléctrico sea mica. Determine el tamaño y número de placas necesarias si se dispone de láminas de mica de espesor 3 μm. Las propiedades de la mica son: permitividad dieléctrica relativa: 7; resistencia dieléctrica: $40 \cdot 10^6$ V/m; resistividad: 10^{13} Ω·cm.

10.2 Se dispone de un sensor capacitivo de proximidad cuyo alcance nominal Sn es de 20 mm. Estudie la posibilidad de poder utilizarlo para detectar la presencia de amoníaco detrás de un panel de cristal de 3 mm de espesor. En caso afirmativo, proponga un procedimiento que actuando sobre el potenciómetro de ajuste de sensibilidad del sensor permita tener el mismo margen de error en ambas direcciones. La figura 10.23 muestra la relación entre el alcance nominal y la constante dieléctrica relativa [10-14].

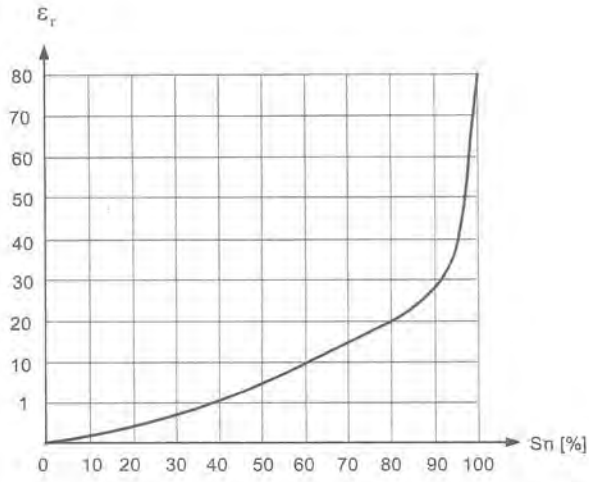


Figura 10.23

10.3 *Empleando el puente de Schering de la figura 10.24, determine:* a) el valor del condensador desconocido C_xR_x, cuando el puente está en equilibrio; b) la tangente de pérdidas de dicho condensador.

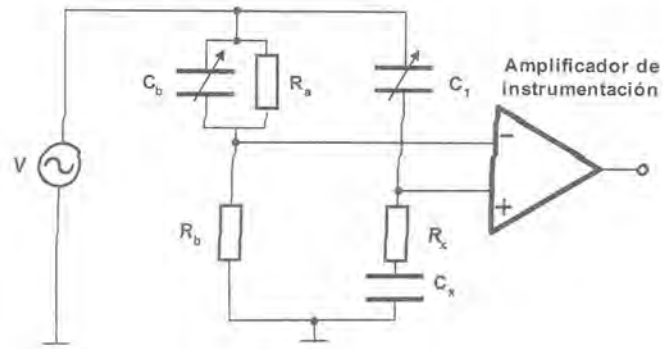


Figura 10.24

10.4 Una fábrica de caramelos dispone de un depósito de azúcar como el que se muestra en la figura 10.26. Se pretende diseñar un sencillo interruptor de nivel capacitivo que genere una señal de alarma cuando el nivel del depósito caiga por debajo de un determinado valor ajustable. Determine: a) la expresión de la capacidad en función de la altura; b) el circuito equivalente que represente todas las capacidades asociadas, indicando la importancia que cada una de ellas presenta; c) diseñe el circuito de medida que genere la señal de alarma empleando el oscilador CMOS ICM7555 [10.5]; d) si el depósito contuviera un liquido conductor, ¿qué consideraciones de funcionamiento cabría realizar?

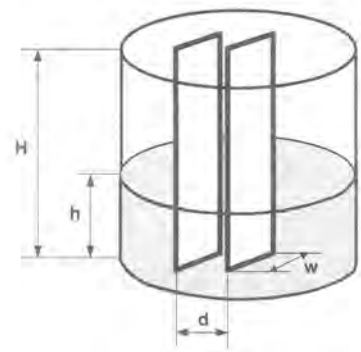


Figura 10.25

10.5 Explicar cómo utilizar el acelerómetro capacitivo de la figura 10.22 para medir la inclinación.

10.6 Se desea conectar el sensor de humedad capacitivo HIH-3610 (tabla 10.4) a un convertidor A/D que admite señales de entrada entre 0 y 5 V. Proponga un posible circuito de acondicionamiento.

11

Sensores inductivos

11.1 Introducción
11.2 Sensores inductivos básicos
11.3 El transformador diferencial lineal (LVDT)
11.4 Otros sensores inductivos

11.1 Introducción

De una forma más o menos equivalente a cómo se usan los potenciómetros y los condensadores como sensores electrónicos cuando la variable a medir afecta a uno de sus parámetros característicos, los dispositivos inductivos pueden también estar afectados por determinadas variables y, por ello, también pueden usarse de una forma similar; lógicamente, los principios de funcionamiento son diferentes y también lo serán los circuitos de medida.

En los sistemas inductivos, el parámetro que se manejará será el valor de la autoinducción L del dispositivo aunque, dadas las posibilidades de transferir señales entre varios devanados, este tipo elementos tendrán otras posibilidades como las derivadas del uso del factor de acoplamiento como variable. A continuación, se estudiarán las posibilidades que presentan estos sistemas. Para ello, comenzaremos por explicar cómo se comporta una bobina genérica y cuáles son los parámetros que la definen.

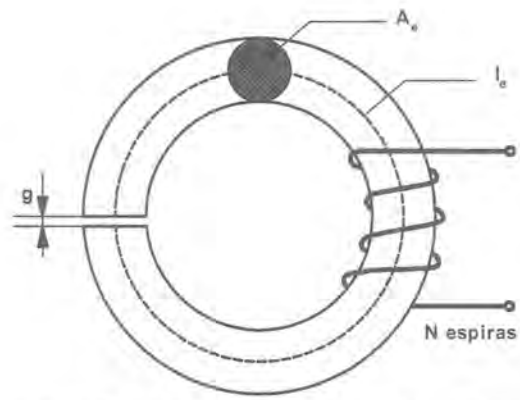


Figura 11.1. Toroide equivalente a un núcleo magnético.

En general, todo conductor tiene un equivalente inductivo que se manifiesta en el valor de su autoinducción; en muchas ocasiones, ese es un factor no deseado que suele ocasionar problemas en algunos sistemas. En otras ocasiones, como la que nos ocupa, lo que se pretende no es eliminar o reducir ese parámetro sino alcanzar un determinado valor para lo que se suele acudir a devanar el conductor formando una bobina, con o sin presencia de un núcleo de material magnético que refuerce el efecto.

Supóngase que se tiene un núcleo magnético al que se han arrollado un número N de espiras formando un devanado; pues bien, todo núcleo magnético de tipo cerrado puede asimilarse a un toroide equivalente de longitud efectiva l_e, área efectiva A_e y entrehierro de longitud g, según se muestra en la figura 11.1.

Si se considera que μ_0 es la permeabilidad magnética del aire y μ la permeabilidad del material magnético empleado, se tiene que:

$$\oint \vec{H} \cdot d\vec{l} = \sum NI \tag{11.1}$$

siendo I la corriente que circula por el devanado. Esta expresión se traduce en:

$$H_1 l_1 + H_2 l_2 = NI \tag{11.2}$$

donde l_1 y l_2 son las longitudes de los caminos correspondientes al aire y al propio núcleo magnético y H_1 y H_2 son los valores respectivos de las intensidades de campo magnético en ambos caminos. Como:

$$B = \mu H$$

sustituyendo en (11.2), se tiene que:

$$\frac{B}{\mu_0} g + \frac{B}{\mu} l_e = NI \tag{11.3}$$

Por otra parte, el flujo magnético es:

$$\phi = \oiint B dS$$

Para este caso, la expresión anterior se convierte en:

$$B = \frac{\varphi}{A_e}$$

con lo que, sustituyendo en (11.3), se obtiene la expresión:

$$\frac{\phi}{A_e}\left(\frac{g}{\mu_0} + \frac{l_e}{\mu}\right) = NI \tag{11.4}$$

Finalmente, se tiene que:

$$L\frac{di}{dt} = N\frac{d\phi}{dt}$$

es decir,

$$LI = N\phi$$

que, junto a (11.4), proporciona el valor de L:

$$L = \frac{A_e N^2}{\frac{g}{\mu_0} + \frac{l_e}{\mu}} \tag{11.5}$$

Si, como es normal, el material magnético tiene una permeabilidad relativa alta, mucho mayor que la del aire, es decir, $\mu_0 << \mu$, la expresión (11.5) se traduce en una forma aproximada pero más sencilla:

$$L \approx \frac{\mu_0 A_e N^2}{g} \tag{11.6}$$

De los diferentes factores que configuran las expresiones (11.5) y su forma aproximada, la (11.6), todos aquellos sobre los que pueda influir una magnitud física cualquiera pueden ser tenidos en cuenta como variables intermedias y pueden dar pie a pensar en utilizar este tipo de sistemas como sensores de medida.

11.2 Sensores inductivos básicos

11.2.1 SISTEMAS CON SENSORES INDUCTIVOS

La variación de los parámetros de las expresiones (11.5) y (11.6) permite realizar diversas configuraciones de sensores inductivos cuando una variable externa afecta a cualquiera de sus parámetros; desgraciadamente, no es fácil la variación de algunos de los valores, las variaciones no son lineales o presentan cambios discontinuos o bruscos que los hacen poco útiles para la mayoría de las aplicaciones. Sin embargo, resulta relativamente sencillo utilizar el parámetro "longitud del entrehierro", g para la medida.

En la figura 11.2 se muestra un primer sistema para la medida de desplazamientos basado en el cambio de la longitud del entrehierro.

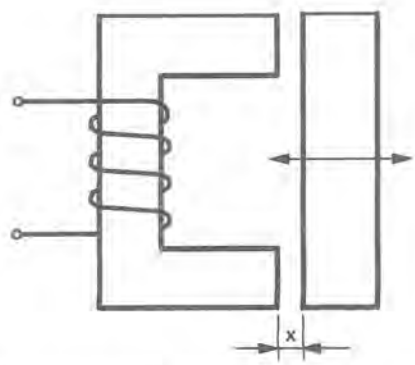


Figura 11.2. El desplazamiento de la pieza afecta al entrehierro x causando variación de la inductancia.

Para este sistema, el valor del entrehierro total, 2x, es la variable a medir, con lo que, aplicando la expresión (11.6):

$$L = \frac{K}{x} \tag{11.7}$$

donde K es una constante que depende del área efectiva y del número de espiras. Este sistema presenta dos limitaciones importantes: en primer lugar, es evidente su falta de linealidad y, en segundo lugar, la validez de las expresiones (11.5) y (11.6) es menor en la medida que el valor del entrehierro sea grande; esto hace que un sistema similar al de la figura 11.2 tenga un campo de aplicación muy reducido y sólo es eficaz en la medida de pequeños espesores de materiales no magnéticos.

Si el núcleo magnético es abierto, el campo de aplicaciones puede crecer aunque las expresiones anteriores no tengan validez. En la figura 11.3 se muestran dos configuraciones para la medida de desplazamientos de los núcleos magnéticos sin la restricción de valor impuesta por el sistema de la figura 11.2.

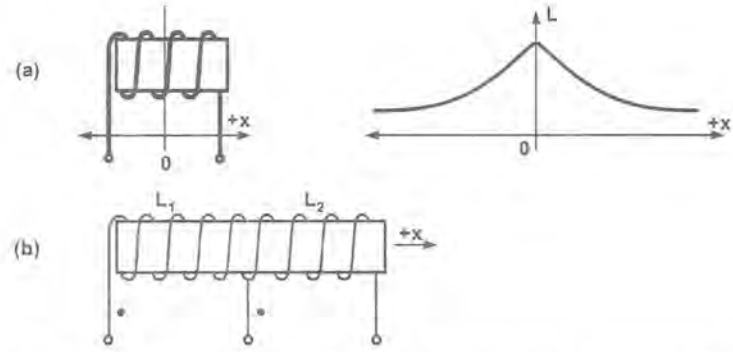


Figura 11.3. Sensores inductivos para medida de desplazamiento y respuesta de uno de ellos.

En la figura 11.3a se muestra la relación que existe entre el valor del desplazamiento, x, y el de L que generalmente debe obtenerse por métodos empíricos y/o de cálculo numérico dada la escasa precisión de las expresiones matemáticas en estos casos. De la observación de la curva de calibración del sistema se desprende la presencia de simetría respecto del eje de ordenadas, lo que equivale a simetría respecto a un plano perpendicular del núcleo que pase por su punto medio. En resumen, se traduce en la imposibilidad de distinguir entre los dos posibles sentidos del desplazamiento. Finalmente, otro factor negativo es la falta de linealidad aunque, bajo este punto de vista, sí que existen algunas zonas aprovechables.

Algunas otras configuraciones para la medida de desplazamientos se tienen en los ejemplos de la figura 11.4 que se pueden usar directamente en una configuración en puente ya que presentan un comportamiento de tipo diferencial, añadiendo L a una de las bobinas al tiempo que se le resta a la otra.

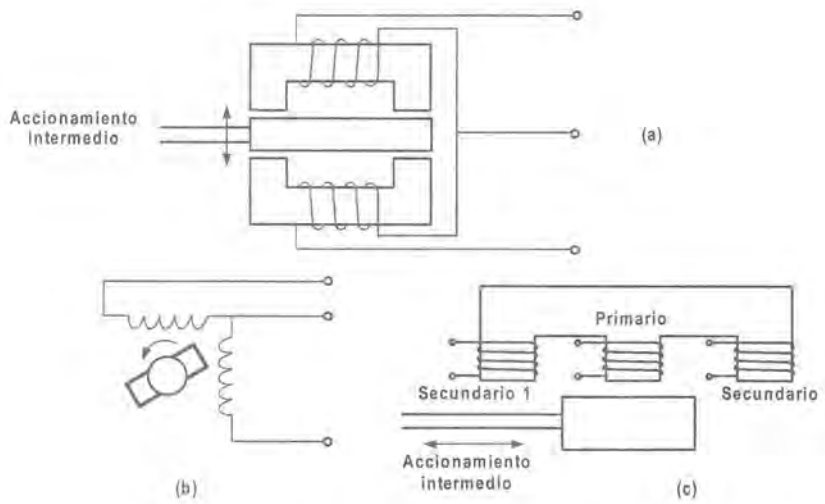


Figura 11.4. Sensores de posición de tipo inductivo: (a) gran sensibilidad y corto alcance; (b) angular; (c) detector de gran alcance.

El uso de configuraciones de este tipo va a permitir detectar el signo del desplazamiento al producirse una inversión en la fase de la señal de salida, según se verá más adelante.

11.2.2 CIRCUITOS DE MEDIDA

Los circuitos de medida para este tipo de sensores son similares a los empleados en los sensores capacitivos; se tratará de una medida en alterna, pudiendo utilizarse cualquiera de las técnicas que se usan con ellos tales como la medida de tensión en un sistema excitado por corriente, medida de corriente en un sistema excitado por tensión, circuitos resonantes o en puente. En todos estos casos, recordemos que la excitación debe ser constante no sólo en el valor de tensión sino también en frecuencia.

En el caso particular de sensores inductivos como los de la figura 11.4a, con una topología con dos arrollamientos y un comportamiento de tipo diferencial, en el que una de las bobinas sufre aumento en el valor de su L y la otra, una disminución, el circuito de medida más adecuado es la configuración en puente como el de la figura 11.5, en el que la tensión de salida aparece entre los puntos A y B.

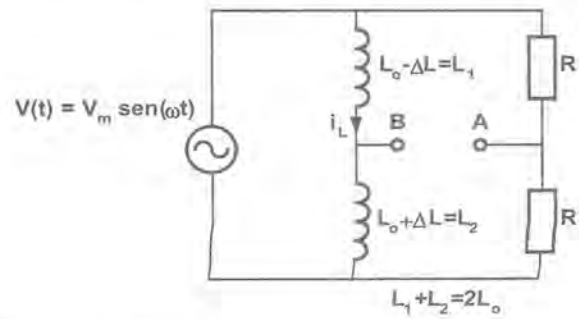


Figura 11.5. Circuito de medida para bobinas diferenciales.

Para este circuito, se tiene que:

$$i_L(t) = -\frac{V_m}{2\omega L_0}\cos(\omega \cdot t)$$

por lo que:

$$V_B(t) = (L_0 + \Delta L)\frac{di_L}{dt} = (L_0 + \Delta L)\frac{V_m}{2L_0}\text{sen}(\omega \cdot t) = \frac{1}{2}V_m\text{sen}(\omega \cdot t) + \frac{\Delta L}{2L_0}V_m\text{sen}(\omega \cdot t)$$

Como quiera que:

$$V_A(t) = \frac{1}{2}V_m\text{sen}(\omega \cdot t)$$

se tiene que:

$$V_B - V_A = \frac{\Delta L}{2L_0}V_m\text{sen}(\omega \cdot t) \qquad (11.8)$$

De esta manera, la magnitud del cambio se puede conocer por medio del valor máximo de la señal de tensión entre los puntos A y B. Pero la expresión (11.8) sólo proporciona el valor absoluto del cambio de L

pero no su sentido. Sin embargo, también es posible obtenerlo midiendo el ángulo de desfase respecto de la señal de alimentación $v(t) = V_m \text{sen}(\omega t)$ ya que, cuando el valor de ΔL es positivo, la fase de la señal $V_B - V_A$ es nula pero cuando ese incremento es negativo, la fase es de 180°. Aprovechando este hecho, se puede aplicar el circuito de la figura 11.6 que introduce el signo en el valor de la señal obtenida de tal manera que la salida es positiva cuando ΔL sea positivo y, negativa cuando el incremento sea negativo.

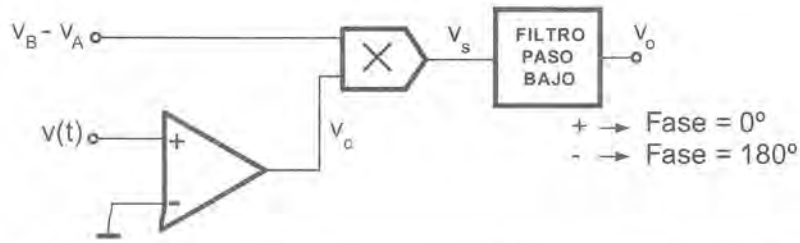


Figura 11.6. Detector de signo del desplazamiento basado en la detección de fase.

El funcionamiento básico es muy simple: la señal v(t) se compara con cero para obtener una señal cuadrada v_C a la salida del comparador; esta señal se multiplica por la señal $v_B - v_A$ obteniéndose un valor siempre positivo o siempre negativo según coincidan o no las fases. Este valor v_S presenta un rizado importante que puede eliminarse mediante un filtro paso bajo obteniendo su valor medio, v_m (véase el capítulo 4). En la figura 11.7 pueden verse las formas de onda teóricas del circuito de la figura 11.6.

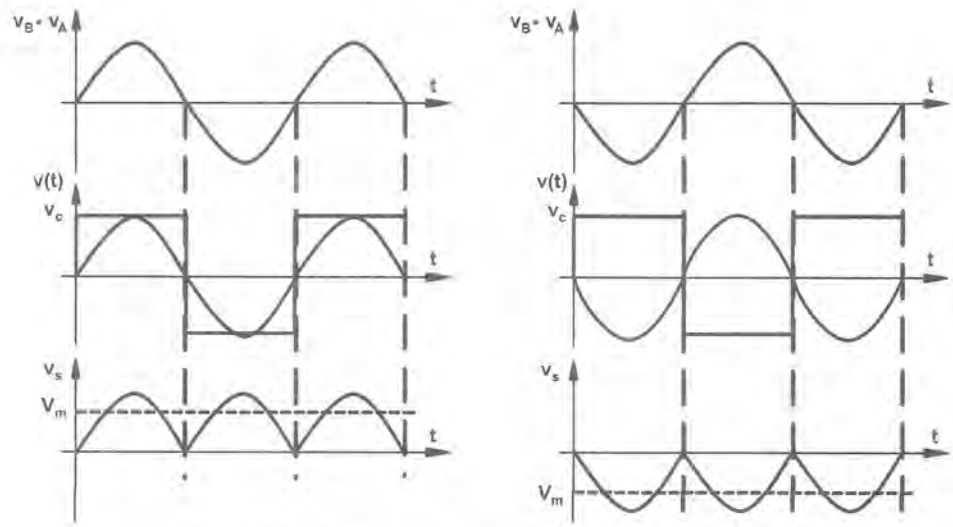


Figura 11.7. Formas de onda del detector de signo de desplazamiento.

La funcionalidad que proporciona este circuito puede obtenerse mediante otras topologías que empleen multiplexadores analógicos que permitan seleccionar entre las dos posibilidades de señal, positiva o negativa.

11.2.3 SISTEMAS DE MEDIDA CON SENSORES INDUCTIVOS

El inconveniente fundamental de todos los sistemas inductivos vistos hasta ahora está en la no linealidad de la variable de salida del sistema frente a la distancia e, incluso, la dificultad de obtener expresiones explícitas para relacionar L con el desplazamiento. Esto obliga a la obtención de curvas de calibración experimentales y, como consecuencia, la utilización de sistemas de almacenamiento y procesamiento digital de la información.

Finalmente, hay que destacar la aplicación de una configuración muy especial basada en un dispositivo de tipo magnético pero que no busca la detección del valor de L, el LVDT, que se verá a continuación.

11.3 El transformador diferencial lineal (LVDT)

11.3.1 DESCRIPCIÓN DE FUNCIONAMIENTO

El transformador diferencial lineal o LVDT como se le suele conocer más a menudo (de las iniciales anglosajonas *Linear Variable Differencial Transformer*) es uno de los dispositivos más empleados en la medida de desplazamientos ya que presenta unas buenas características de linealidad y de sensibilidad que se comentarán a continuación. Muchos sistemas en los que la variable a medir es un desplazamiento y que no precisan la medida a distancia utilizan como sensor electrónico final un LVDT.

La estructura de un LVDT es la que se muestra en la figura 11.8, donde se pueden ver sus tres arrrollamientos, el del medio (de N_1 espiras) actúa como primario del transformador y, los otros (de N_2 espiras cada uno), como secundarios.

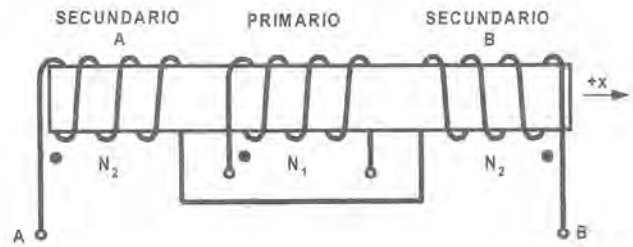


Figura 11.8. Configuración básica de un LVDT.

Si los secundarios se unen como se indica en la figura 11.8 (teniendo en cuenta sus terminales correspondientes) y se introduce una señal senoidal en el primario del transformador cuyo valor máximo sea V_1, la máxima tensión que aparece en la salida situada entre los terminales A y B es la siguiente:

$$V_{AB} = \frac{N_2}{N_1}\left(\frac{d\phi_{2a}}{d\phi_1} - \frac{d\phi_{2b}}{d\phi_1}\right)\cdot V_1 \qquad (11.9)$$

donde ϕ_1 es el flujo generado por el primario, ϕ_{2a} es la parte del flujo ϕ_1 concatenado por el secundario A y ϕ_{2b} es la parte del flujo ϕ_1 concatenado por el secundario B. ϕ_{2a} y ϕ_{2b} se modifican según se varíe la posición del núcleo magnético tendiendo a disminuir la parte de flujo que concatena el devanado sobre el que aparece

la menor parte del núcleo. Así, el desplazamiento presenta una influencia directa en el valor de la tensión obtenida. Si se representa este valor, el aspecto obtenido es como el que se muestra en la figura 11.9.

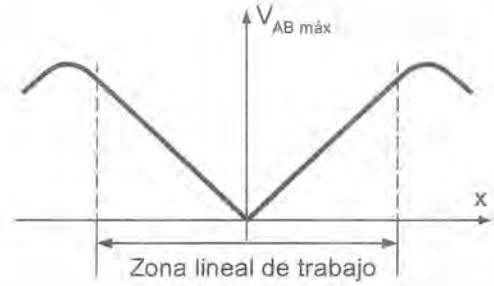


Figura 11.9. Respuesta del LVDT en función del desplazamiento del núcleo.

En este gráfico se aprecia una zona muy lineal —para valores positivos o negativos del desplazamiento x— que se puede utilizar como margen de trabajo del sensor. También puede verse que, igual que en los sistemas de la figura 11.3, no es factible detectar el signo del desplazamiento por la simetría que presenta el sistema.

11.3.2 CIRCUITOS DE MEDIDA

Mediante un circuito similar al presentado en la figura 11.6, es posible detectar el sentido del desplazamiento. No obstante, también se puede plantear una alternativa (válida para ambos casos) en la que se sustituye el multiplicador analógico (dispositivo de elevado coste y no muy buenas prestaciones) por otros dispositivos menos problemáticos como se muestra en la figura 11.10.

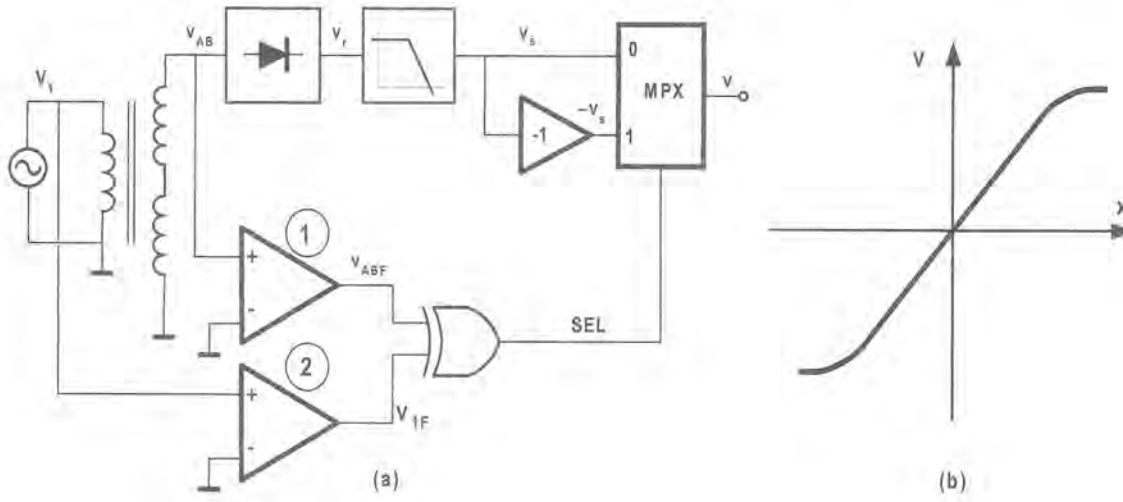


Figura 11.10. (a) Circuito para la medida con signo del desplazamiento del núcleo de un LVDT, y (b) característica de salida en que se observa el efecto del signo del desplazamiento sobre la tensión de salida.

El funcionamiento es bastante sencillo, pese a lo que pueda parecer en un principio. En los comparadores 1 y 2 se obtiene una señal cuadrada con la misma fase que V_{AB} y V_1, respectivamente. Si las fases son iguales, la puerta EXOR proporciona una salida a nivel lógico bajo y se selecciona el canal 0 del multiplexador analógico (MPX); si, por el contrario, las señales están desfasadas 180°, la salida de la puerta estará a nivel lógico alto y el canal seleccionado en MPX es el 1.

Por otro lado, al canal 0 del multiplexador se llevará el valor medio de la señal (para ello se habrá rectificado y filtrado) y, al canal 1, ese mismo valor pero multiplicado por −1, es decir, con el mismo valor absoluto pero signo contrario. En la figura 11.11 pueden verse las distintas formas de onda que se obtienen en el sistema para un desplazamiento positivo (trazo continuo) y para uno negativo (trazo discontinuo), suponiendo que el desplazamiento positivo implique fase 0. En la figura 11.10b, aparece la curva de transferencia entrada (desplazamiento) – salida (tensión) teniendo en cuenta ya el procesamiento de la señal con el circuito de la figura 11.10a. Se puede observar una amplia zona lineal alrededor del punto x = 0 que es susceptible de ser usada para la medida lineal y que le da el nombre a este sensor.

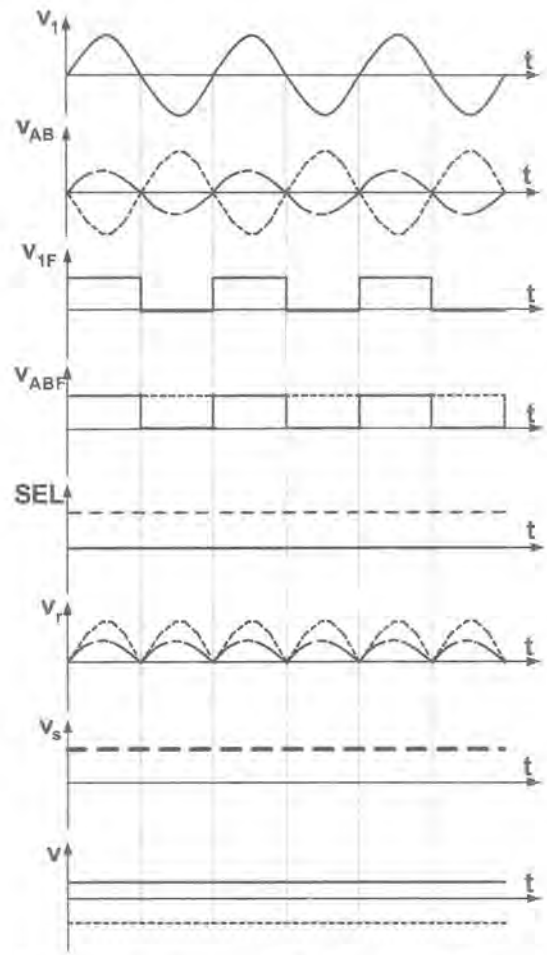


Figura 11.11. Tensiones del circuito de la figura 11.10; como se puede ver, la salida de tensión incluye el signo del desplazamiento.

Es posible, claro está, trabajar fuera de la zona lineal de uso del LVDT y mediante su curva de calibración [11.1] o mediante el uso de un ajuste polinómico (suele ser suficiente uno de grado 2), lo que implica, en general, la inclusión de algún sistema de cálculo digital. En estas condiciones, es posible ampliar el margen de medida hasta el doble del lineal pero el precio a pagar es la complicación del circuito y el cambio de la sensibilidad a lo largo de toda la zona de uso.

El circuito de la figura 11.10 es sólo una primera aproximación a la realidad ya que presenta una serie de problemas que será preciso resolver para obtener un óptimo comportamiento. En concreto habrá que tener en cuenta los siguientes factores:

a) El bloque de rectificación debe ser de precisión y de doble onda para garantizar la ausencia de la zona muerta producida por la tensión de codo de los diodos y un bajo rizado de la tensión de salida.

b) Los comparadores deben proporcionar niveles de salida aptos para que sean admitidos por las entradas de la puerta EXOR. También debería situarse un pequeño filtro (posiblemente sólo una red RC) a la salida de la puerta para evitar picos en los instantes en que el núcleo del LVDT está centrado y la tensión de salida es próxima a cero.

c) El multiplexador analógico debe permitir el paso de señales positivas y negativas.

Existen, además otros condicionantes propios de la aplicación tales como la frecuencia de la señal, el tipo de filtrado, etc. que condicionarán el comportamiento dinámico del conjunto y los requisitos exigibles a los amplificadores operacionales que se utilicen.

Los LVDTs son dispositivos muy empleados en aplicaciones industriales, bien sea para medir desplazamientos directos o para la medida de variables que produzcan un desplazamiento como variable intermedia. También hay sistemas basados en LVDT como en el caso de accionamientos realimentados; un ejemplo de estos es la servo-válvula en la que el sistema de control proporciona el dato de la posición a la servo-válvula y esta se mueve hasta la posición especificada usando como sensor de posición un LVDT [11.7].

Otra posible solución, más práctica habitualmente que desarrollar un circuito discreto, es acudir a soluciones comerciales dedicadas —siempre que haya algún fabricante que las ofrezca— ya que suelen ahorrar tiempo de desarrollo y reducir, en consecuencia, los costes. Para el caso de los LVDTs, varios fabricantes ofrecen circuitos integrados monolíticos para esta aplicación como el 5521 de Philips [11.4] o el AD698 de Analog Devices [11.2]. En ambos casos, la solución precisa pocos componentes externos ya que ambos incluyen el sistema de excitación y los sistemas de demodulación de la señal de salida [11.3].

11.3.3 CONSIDERACIONES PRÁCTICAS EN EL DISEÑO DE CIRCUITOS CON LVDT

El funcionamiento del LVDT con alguno de los circuitos vistos hasta ahora parece muy sencillo; sin embargo, hay que tener unas ciertas precauciones para usarlo con vistas a no ver defraudadas nuestras expectativas iniciales.

Señal de excitación del primario

¿Por qué no utilizar una forma de onda cuadrada? Es más sencilla de obtener y eso simplificaría el diseño. El problema está en que los factores que caracterizan la respuesta del sensor: linealidad y sensibilidad son dependientes de la frecuencia y, por ello, la onda cuadrada introduciría efectos de incertidumbre en la medida. También se producirían problemas derivados de los cambios introducidos por la temperatura y se complicaría el diseño de las etapas de demodulación y filtrado que se encontrarían con varias frecuencias a procesar.

Tampoco es necesaria una tensión de excitación del primario perfectamente senoidal —sería deseable pero resultaría costosa— y suele aceptarse como buena una señal con una distorsión baja: un valor de THD menor del 2 ó 3% se puede considerar como buena [11.6].

Por otro lado, podríamos preguntarnos por la frecuencia deseable para la señal de excitación. Como en todos los casos de ingeniería, habrá que llegar a una solución de compromiso puesto que habrá factores que se vean favorecidos por un ascenso en la frecuencia y otros, por el descenso. Las frecuencias altas serían poco deseables puesto que complicarían su tratamiento (necesitaría circuitos más rápidos) y necesitaría un diseño, en general, más cuidadoso; sin embargo, una frecuencia elevada simplifica el filtro de la salida (figura 11.10). En efecto, en un sistema de medida de la posición, tenemos dos señales diferentes: la señal de excitación que actúa como portadora y la señal de posición que actúa como moduladora. Esta última señal es la que debe ser demodulada para recuperar la información y este proceso se facilita en la medida en que la portadora esté más alejada de ella en frecuencia. Por tanto, o bien tenemos una portadora de alta frecuencia o sólo podremos recuperar señales de baja frecuencia. En la misma línea, si la posición cambia rápidamente, será preciso utilizar una portadora de mayor frecuencia.

En las típicas condiciones de trabajo, este tipo de sensores no se utilizan para medir vibraciones ni movimientos rápidos debido a la elevada masa de núcleo y vástago que afectaría gravemente a la variable a medir, por lo que cabe pensar en el LVDT como en un sensor para medidas estáticas o pseudo-estáticas. De esta manera, frecuencias de trabajo de unos pocos kHz son más que suficientes.

Ajuste del cero y del alcance

Para garantizar la independencia de los ajustes de cero y valor máximo (alcance) del sistema se puede incluir una etapa adicional en el sistema de amplificación y *offset* situada detrás del multiplexador e inmediatamente antes del filtro. Un pequeño amplificador con ganancia ajustable y nivel de continua variable puede fijar el valor de salida para el punto central (0 V) y para uno de los extremos de medida ($+V_{máx}$ o $-V_{máx}$); el sistema a utilizar puede ser como el de la figura 11.12. Obsérvese que dada la ganancia negativa de la etapa de amplificación, los signos de la salida estarán cambiados respecto de los de la figura 11.10.

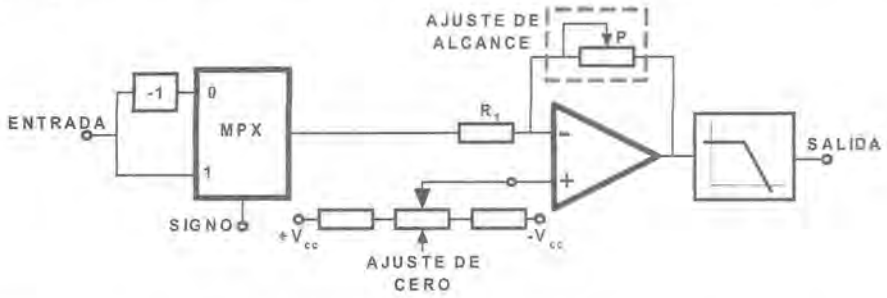


Figura 11.12. Circuito para ajustar el alcance y el cero en un sistema basado en LVDT.

También sería posible no incluir más amplificadores operacionales en el sistema y aprovechar las posibilidades de las etapas ya disponibles: por ejemplo, podría realizarse un ajuste de *offset* en la etapa de filtrado que resulta relativamente fácil. Algunos de los circuitos comerciales que se comentaron antes permiten determinados ajustes para fijar el cero y el extremo.

Independencia de la excitación

Según se comentó antes, la estabilidad de la fuente de excitación del primario resulta crucial para el funcionamiento del LVDT ya que la tensión de salida no sólo depende de la posición del núcleo sino también depende de esa señal.

Para conseguir que la información que aparezca en la salida sea una buena representación de la entrada (el desplazamiento), debería eliminarse la influencia de cualquier perturbación de la señal de excitación lo que se consigue mediante un diseño muy optimizado de la fuente de excitación o mediante la eliminación de su influencia sobre la salida. La primera alternativa resulta compleja y nunca se podrá hacer la salida independiente de las interferencias de cualquier tipo, provengan de la fuente o, incluso, del propio LVDT. En la figura 11.13 se pueden observar otras alternativas que permiten hacer que la señal de salida sólo dependa del desplazamiento del núcleo del LVDT.

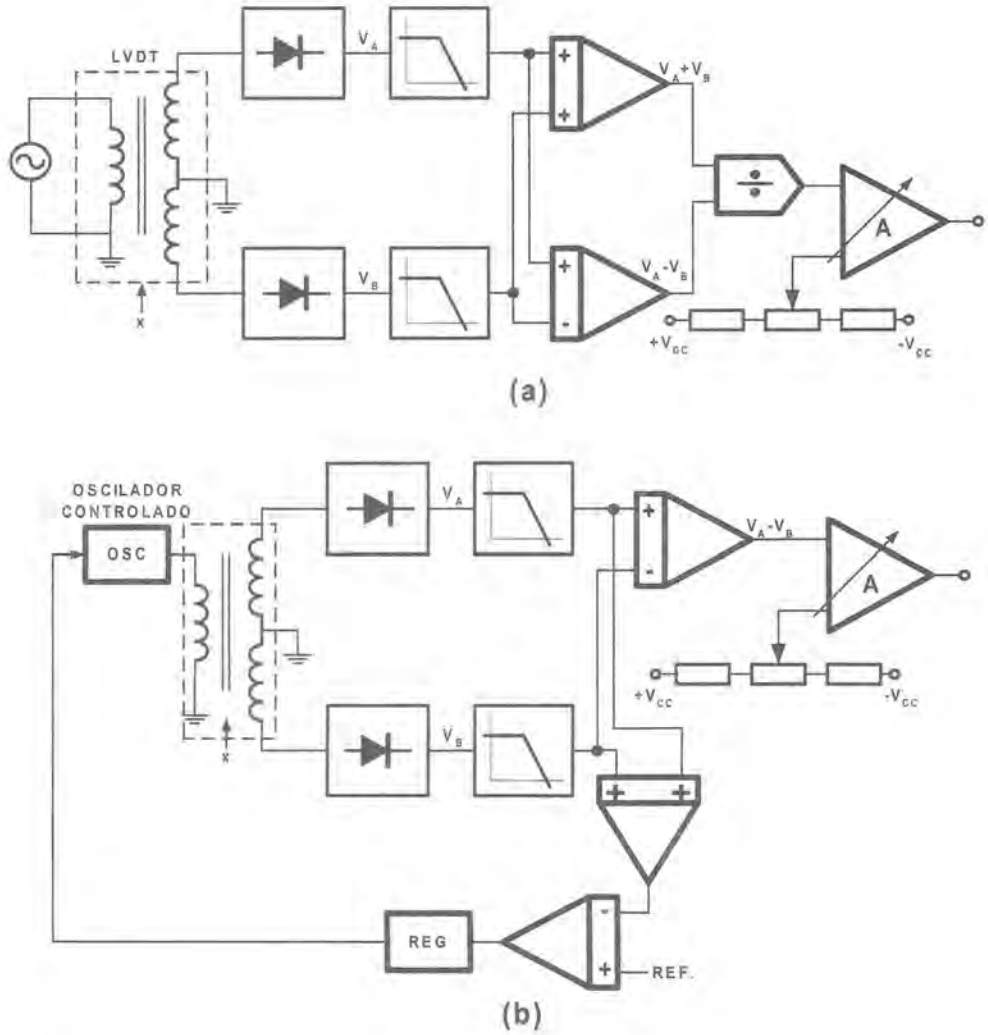


Figura 11.13. Circuitos para independizar la salida de un LVDT de la excitación: (a) mediante medida de la excitación y, (b) mediante un oscilador realimentado.

El sistema de la figura 11:13a está midiendo permanentemente la señal de excitación a través de la suma de las señales inducidas en los secundarios (V_A+V_B) y está dividiendo la salida por este factor con lo que se eliminan las influencias de cambios tanto en la fuente como en los coeficientes de acoplamiento primario-secundarios. Esta es la técnica seguida por el circuito AD598 de Analog Devices. Por el contrario, en la

figura 11.13b se muestra un oscilador realimentado que controla el valor de su señal de salida gracias a la medida de la señal que llega al conjunto de los secundarios del LVDT.

EJERCICIO 11.1

El devanado primario de un sensor LVDT está excitado con una fuente de alimentación senoidal de 5 kHz y dispone de un circuito de tratamiento similar al de la figura 11.10. La tensión de la fuente es tal que la tensión entre los puntos A y B es de 5,000 V de pico cuando está en cualquiera de los extremos del margen lineal (±10,00 mm). Si a la salida hay un filtro de Butterworth de tercer orden con frecuencia de corte de 500 Hz, se pide:

a) ¿Cuál es la sensibilidad del conjunto?

b) ¿Qué rizado aparece en la salida cuando el sensor está en el fondo de escala?

c) ¿Qué error se introduce en la señal de salida cuando el sensor debe medir un movimiento oscilatorio de 500 Hz de frecuencia?

SOLUCIÓN:

a) Cuando el núcleo del LVDT está en uno de los extremos, siguiendo el esquema eléctrico de la etapa de tratamiento de la figura 11.10, a la salida del bloque de rectificación de onda completa y suponiendo que los amplificadores están correctamente elegidos para trabajar en estas condiciones, tendremos una señal rectificada como la de la figura 11.14a.

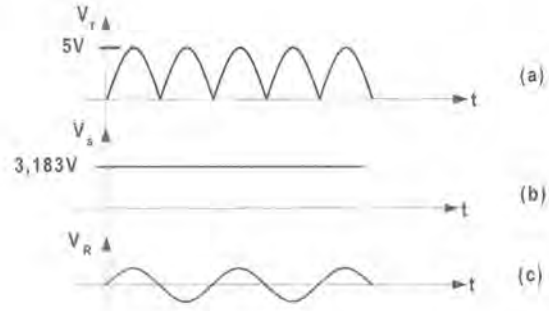


Figura 11.14

Esta señal tiene un valor medio V_S de:

$$V_S = \frac{2}{\pi} V_{máx} \Rightarrow V_S = 3{,}183 \text{ V}$$

Si consideramos que el resto de las etapas del circuito no aportan ganancia alguna, la salida que tendríamos sería de ±3,183 V para una posición de entrada de ±10,00 mm. Por tanto, la sensibilidad del sistema resultaría ser S = 0,3183 V/mm.

b) El desarrollo en serie de Fourier de la onda rectificada permite obtener las distintas señales que la integran. Los armónicos se encuentran a frecuencias múltiplos par de la señal de excitación y serán de valor V_{Rm}:

$$V_{Rm} = \frac{\sqrt{2}}{4n^2 - 1} V_{máx}$$

Así pues, los diversos armónicos se producirán a frecuencias de 10, 20, 30 kHz y así sucesivamente. Sus valores serán:

$$\begin{aligned} n = 1\ (10\ kHz) &\rightarrow V_{R1} = 2{,}357\ V \\ n = 2\ (20\ kHz) &\rightarrow V_{R2} = 0{,}4714\ V \\ n = 3\ (30\ kHz) &\rightarrow V_{R3} = 0{,}2020\ V \end{aligned}$$

Al pasar por el filtro de salida, las atenuaciones producidas serán mayores a medida que lo sea la frecuencia de cada señal; así, tendremos que la señal de 10 kHz se atenuará en 78 dB (7943 veces), la de 20 kHz en 96 dB (63010 veces) y la de 30 kHz, en 114 dB (501200 veces).

Como se puede observar, las señales quedan muy atenuadas ya que los valores finales son:

$$\begin{aligned} n = 1\ (10\ kHz) &\rightarrow V_{R1} = 0{,}2967\ mV \\ n = 2\ (20\ kHz) &\rightarrow V_{R2} = 7{,}481\ \mu V \\ n = 3\ (30\ kHz) &\rightarrow V_{R3} = 0{,}4030\ \mu V \end{aligned}$$

El rizado de la tensión de salida será la suma cuadrática de los rizados a las diversas frecuencias que no se va a diferenciar mucho del valor a 10 kHz:

$$V_R = \sqrt{\sum_{n=1}^{\infty} V_m^2} \Rightarrow V_R = \sqrt{0{,}2967^2 + 0{,}007481^2 + 0{,}0004030^2} = 0{,}2968\ mV$$

Es decir, no se cometería demasiado error si sólo se considera el primer armónico. Esto es cierto cuando el filtro sea suficientemente selectivo en relación con la frecuencia a eliminar como suele ser habitual si el sistema está correctamente diseñado.

c) Si la señal de entrada no es continua, sino que presenta oscilaciones, estas deberán atravesar el filtro por lo que se verán modificadas por él. En el caso de la frecuencia de 500 Hz en la señal de entrada, coincide con la frecuencia de corte del filtro de Butterworth; así pues, la caída de señal a la frecuencia de corte resulta ser de 3 dB por lo que la señal sería amortiguada en un factor 1,4142. El error en la salida es muy grande: si consideramos una oscilación en todo el margen de medida (desde –10 mm hasta +10 mm) la señal teórica debería variar desde –3,183 hasta +3,183 V pero, por acción del filtro, la señal de salida oscilará desde –2,251 V hasta +2,251 V ¡con un error superior al 40%!

Si se pretendiese trabajar con señales de estas frecuencias en la entrada, el filtro debería ser retocado. Dejamos para el lector tomar las medidas oportunas para ello. Como sugerencia, se puede alejar la frecuencia de corte hasta una década por encima de la señal de entrada e incrementar el orden del filtro de salida (para reducir el rizado que si no, sería alto).

11.4 Otros sensores inductivos

Cuando se habla de sensores inductivos no se suele pensar en alguno de los dispositivos vistos hasta este momento que se suelen identificar por su nombre directo. Por el contrario, vienen a la mente sensores de proximidad inductivos. Este tipo de dispositivos son muy habituales en las plantas industriales ya que presentan características muy interesantes en la medida de proximidad y en la detección de objetos metálicos.

El principio de funcionamiento de este tipo de sistemas es muy simple: consiste en un oscilador cuya salida se aplica a una bobina de núcleo abierto capaz de generar un campo electromagnético en sus proximidades; la presencia de objetos metálicos en la zona modificaría el campo y se manifestaría algún cambio en las magnitudes eléctricas de la bobina. Los cambios pueden detectarse y conseguir, así, saber si existe o no un objeto metálico dentro del radio de acción del sistema.

Un caso particular de este sistema y bastante empleado es el principio ECKO (*Eddy Current Killed Oscillator*) [11.5] que se aplica en un sistema como el de la figura 11.15.

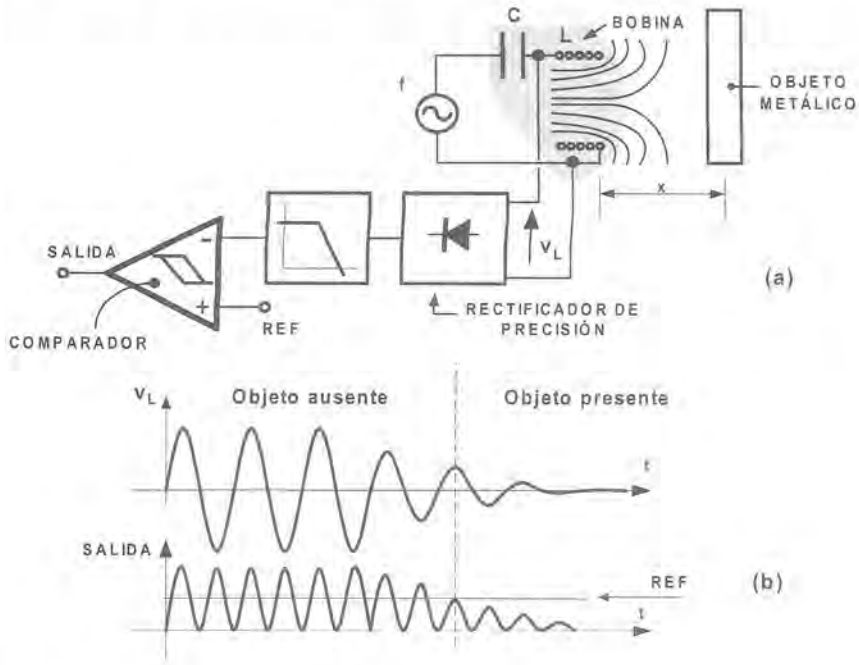


Figura 11.15. Sensor ECKO: (a) circuito básico; (b) formas de onda en el sensor y en la salida.

La fuente estabilizada de alterna produce una señal a frecuencia f a la que está sintonizado el circuito LC, cuya bobina es de núcleo abierto y produce campo magnético en las proximidades. En presencia de un objeto metálico, las corrientes de Eddy que se producen en el objeto provocan una carga adicional en la bobina, carga que es mayor en la medida que el objeto está más próximo. Esta carga produce una caída de la señal en la propia bobina que puede detectarse midiendo la tensión en ella y comparándola con una referencia. Por debajo de esa referencia, se considerará que hay un objeto metálico próximo y, por encima, se considerará que no (figura 11.15b).

Este tipo de sistema suele tener un aspecto como el de la figura 11.16, constituyendo un bloque monolítico con una salida capaz de manejar una carga del tipo de un relé. Existen muchas versiones de este tipo de dispositivo en función del alcance máximo de detección (desde menos de un milímetro hasta pocas decenas de centímetros en función del tipo y tamaño del objeto a detectar), su alimentación (continua o alterna y de diversos valores) y su conexionado exterior.

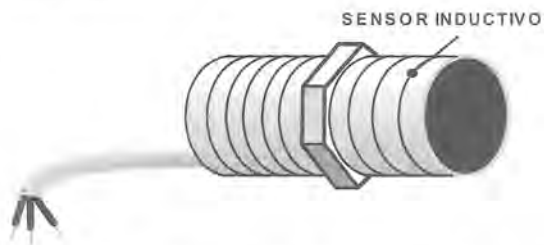


Figura 11.16. Aspecto de un típico detector de proximidad inductivo. El cuerpo cilíndrico roscado permite una fácil colocación e intercambiabilidad en aplicaciones industriales.

En la figura 11.17 se muestran algunas de las formas típicas de conexión de la carga y de alimentación continua de estos dispositivos que suelen ser comunes a la mayoría de los fabricantes.

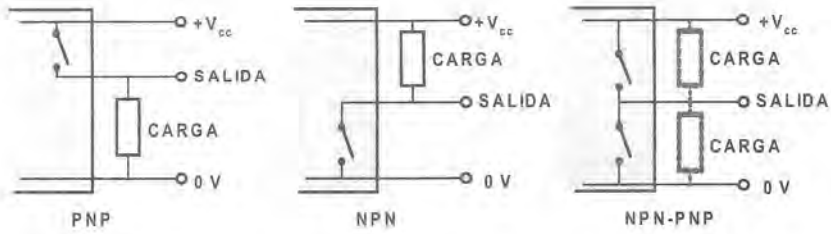


Figura 11.17. Salidas típicas de detectores de proximidad de tipo inductivo con tres hilos de conexión. Obsérvese que la colocación de la carga depende del tipo de salida.

El uso de estos sensores es realmente muy sencillo pero, para optimizar su aplicación, será conveniente tener en cuenta unas cuantas circunstancias que pueden afectar a la detección de objetos:

a) La presencia de una bobina apantallada o no apantallada modifica la distribución del campo, ocasionando un cambio en la distancia de detección del sistema. En la figura 11.18a se muestra el campo producido por una bobina apantallada, mientras que en la figura 11.18b aparece el que provoca una no apantallada.

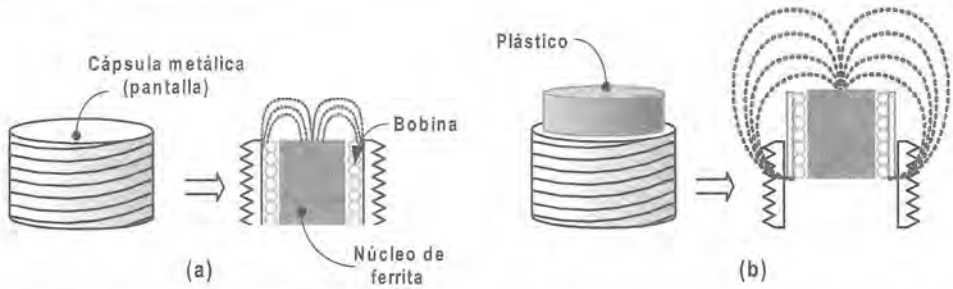


Figura 11.18. (a) Bobina apantallada, y (b) no apantallada en los sensores de proximidad inductivos.

b) El tamaño del objeto resulta, obviamente importante ya que el efecto producido será mayor en la medida en que lo sea el objeto a detectar. O, lo que es lo mismo, un objeto más grande se detectará a una distancia más grande que otro objeto más pequeño. Los fabricantes suelen proporcionar un tamaño de objeto "estándar" y proporcionan los datos de alcance de sus sensores para este objeto. Como guía para saber cómo se modifica el alcance en función del tamaño del objeto, se puede utilizar la tabla 11.1 [11.5].

S_{obj}/S_{st}	Factor de corrección	
	Bobina apantallada	Bobina no apantallada
0,25	0,56	0,50
0,50	0,83	0,73
0,75	0,92	0,90
1,00	1,00	1,00

Tabla 11.1. Factores de corrección para la distancia de detección en sensores de proximidad inductivos con y sin bobina apantallada.

Donde, S_{obj} es la superficie del objeto y S_{st} es la superficie del objeto estándar. Como se puede observar, la bobina apantallada presenta una menor influencia del tamaño del objeto a detectar.

c) El material del objeto es importante porque las corrientes inducidas pueden ser más o menos grandes en función de él y, como consecuencia, lo será su influencia sobre el circuito resonante. En la tabla 11.2 se indican una serie de factores de corrección de la distancia de alcance en función del material [11.5].

Material	Factor de corrección	
	Bobina apantallada	Bobina no apantallada
Cobre	0,30	0,40
Aluminio	0,35	0,45
Latón	0,40	0,50
Acero inoxidable 300	0,70	0,80
Lámina de aluminio	0,90	1,00
Acero	1,00	1,00

Tabla 11.2. Factores del alcance en función de los materiales detectados.

Finalmente, las condiciones de uso del sensor inductivo tales como la tensión de alimentación o la temperatura ambiente a la que esté sometido afectarán ligeramente a la distancia de detección y, en algunos casos habrá que efectuar una corrección más en función de las condiciones de uso. En este sentido, se puede definir una distancia nominal de detección (D_n) que se fija a temperatura estándar (25°C) y a la alimentación nominal y una distancia mínima garantizada que suele ser alrededor de un 20% menos que D_n y que se asegura para todos los dispositivos en todas las condiciones de trabajo permisibles. La utilización de estos factores de corrección puede modificar la selección de la distancia de trabajo para una aplicación concreta.

EJERCICIO 11.2

Se dispone de un sensor inductivo de proximidad de bobina apantallada con un alcance nominal de 20 mm para objetos de acero de superficie 2 cm^2. Si se pretendiese utilizar este dispositivo para detectar objetos de acero inoxidable cuadrados de 12,5 mm de lado a una distancia de 10 mm, ¿serviría este sensor? Utilícense los datos de las tablas 11.1 y 11.2.

SOLUCIÓN:

En primer lugar, tengamos en cuenta la reducción de alcance por la diferencia de superficie. La superficie real es de 12,5x12,5 = 156 mm^2 lo que supone que el parámetro S_{obj}/S_{st} es 0,78. se podría interpolar con los valores de la tabla 11.1 pero, como la medida se va a efectuar es de proximidad (una medida todo-nada), no merece la pena hacerlo y se puede tomar el valor de que corresponde a S_{obj}/S_{st} = 0,75 que, además, es más conservador: la reducción del alcance es de 0,92 lo que significa que el alcance con un objeto de acero de este tamaño será de:

$$\text{Alcance}_{obj} = 0{,}92 \times 20 = 18 \text{ mm}$$

El objeto a detectar es acero inoxidable (suponemos que del tipo del que se indica en la tabla 11.2) lo que significa una reducción del alcance de 0,70 luego el alcance para objetos de este tamaño y material será de:

$$\text{Alcance}_{obj\text{-}inox} = 0{,}70 \times 18 = 13 \text{ mm}$$

Es decir, el alcance nominal del sensor es suficiente para detectar los objetos a este tamaño. Sin embargo, si se consideran los cambios de tensión de alimentación y de temperatura ambiente, deberíamos tener en cuenta modificaciones del alcance de hasta un 20% con lo que el alcance real en las peores condiciones sería un 20% menor, es decir, muy poco más de 10 mm.

Como quiera que hemos optado por todos los parámetros conservadores podemos concluir que el sensor sería suficiente para esta aplicación. Alguien aún más conservador podría pensar que sería mejor opción un sensor con un poco más de alcance.

Resumen

Los sensores de tipo inductivo se basan en el efecto que las diversas variables físicas ocasionan o pueden ocasionar sobre los parámetros que afectan a los valores de auto-inductancia o de acoplamiento magnético de las bobinas.

El primer gran grupo de sensores consiste en la medida de desplazamiento/posición de todo o parte del núcleo magnético modificando el entrehierro o el coeficiente de acoplamiento. Las medidas de este tipo no tienen demasiada linealidad y presentan un alcance muy limitado excepto en el caso del LVDT.

El LVDT o transformador diferencial lineal es uno de los sensores magnéticos más empleados para la medida de posición con bastante precisión aunque requiere un circuito de tratamiento relativamente complejo ya que no es evidente la forma de detectar el signo del desplazamiento.

El sensor magnético de proximidad de tipo inductivo es uno de los más empleados en la detección de proximidad. Su funcionamiento se basa en las modificaciones sobre la distribución del campo magnético ocasionadas por la presencia de objetos metálicos y en el correspondiente cambio en el valor de la inductancia de la bobina que genera el campo.

El diseño con sensores de proximidad inductivos es muy simple aunque deben tenerse en cuenta algunas consideraciones acerca de los materiales a detectar, de la configuración de la bobina del sensor y de la cercanía de otros sensores similares.

Problemas propuestos

11.1 Indique los cambios a realizar sobre el sistema propuesto en el ejercicio 11.1 si la señal de excitación es una onda senoidal del doble de frecuencia.

11.2 ¿Qué ventajas presentaría el uso de ondas cuadradas para excitar el primario de un LVDT? ¿Qué implicaciones de diseño del circuito de acondicionamiento tendría?

11.3 Empleando un NE5521 de Philips (las características se pueden consultar en el CD adjunto) diséñese un sistema semejante al propuesto en el ejercicio 11.1.

11.4 Repítase el problema anterior con el AD698 de Analog Devices (hoja de datos en el CD adjunto).

11.5 A raíz de los resultados de los dos problemas anteriores y de la solución del ejercicio 11.1, ¿cuál de las opciones le parece más apropiada para emplearla en un sistema real?

11.6 Proponga circuitos de acondicionamiento para los detectores de posición de la figura 11.4.

11.7 Empleando las tablas 11.1 y 11.2, indique el alcance de un detector de proximidad inductivo de bobina apantallada con distancia nominal de alcance de 20 mm con un objeto de 10 cm^2 de superficie cuando intenta detectar cuadrados de aluminio de 2 cm de lado.

11.8 En relación con el problema anterior si el detector se encuentra con objetos de acero cuadrados de 1 cm de lado, ¿a qué distancia los detectaría? Compare el resultado con el obtenido en el problema anterior e indique si esto constituye o no un problema.

11.9 Repita el problema anterior con cuadrados de acero del mismo tamaño que los de aluminio.

12

Sensores electromagnéticos

12.1 Introducción
12.2 Sensores electromagnéticos lineales
12.3 Sensores electromagnéticos rotativos

12.1 Introducción

Los sensores electromagnéticos se utilizan desde hace mucho tiempo para medir tensiones y corrientes variables con el tiempo, potencia eléctrica, velocidades lineales y angulares, así como posición angular.

Están basados en la ley de Faraday-Henry de la inducción electromagnética, que señala que si el flujo magnético Φ asociado con un conductor cambia con el tiempo t, en el conductor se induce una fuerza electromotriz e (f.e.m) relacionada con el cambio del flujo mediante la expresión:

$$e = -\frac{d\Phi}{dt} \tag{12.1}$$

El signo menos significa que la polaridad de la tensión inducida haría circular por el circuito del conductor una corriente cuyo flujo asociado se opondría a los cambios del flujo Φ (Ley de Lenz).

Lo anterior hace comprender que si se produce un desplazamiento relativo entre el conductor o una bobina en una región en la que exista un campo magnético constante, la f.e.m. inducida en el conductor o bobina será dependiente de la velocidad con la que se produzca este desplazamiento, lo que da lugar al fundamento de los sensores electromagnéticos de velocidad lineal y angular.

Si el flujo variable en el tiempo creado por una bobina excitada con una corriente alterna (primario) concatena con otro devanado (secundario), se inducirá en este último una tensión que depende fundamentalmente de la relación entre el número de vueltas de los devanados, de la magnitud de la corriente de excitación, su frecuencia y del acoplamiento magnético existente entre los devanados. Si la posición relativa de los devanados es fija, la tensión del secundario dependerá de la excitación del primario dando lugar a los sensores de

tensión, corriente y potencia (transformadores). Si la posición relativa de los devanados se hace cambiar manteniendo constantes los parámetros de la excitación variable con el tiempo, se producirán cambios en la tensión inducida, lo que fundamenta el principio de funcionamiento de los sensores electromagnéticos de posición.

En este capítulo nos referiremos a los también denominados sensores electromecánicos, es decir, a los sensores electromagnéticos en los que existe movimiento relativo entre sus componentes.

12.2 Sensores electromagnéticos lineales

Un sensor electromagnético para la medida de la velocidad lineal está formado por un devanado fijo dentro del cual se mueve un imán permanente solidario al objeto cuya velocidad lineal se desea medir, tal como se muestra en la figura 12.1.

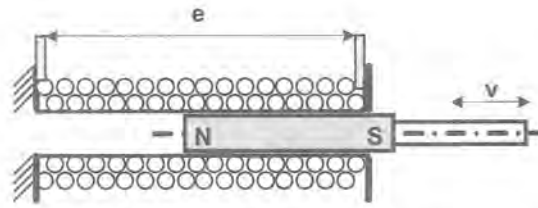


Figura 12.1. Sensor electromagnético de velocidad lineal. Diagrama básico.

En cualquier posición que se encuentre el imán deslizante, si su velocidad lineal es nula, la tensión inducida en el devanado también lo será ya que el flujo magnético del imán que concatena con la bobina no cambia con el tiempo. Si el eje acoplado al imán se desplaza en la dirección axial, en cualquier sentido, a una velocidad **v**, el flujo que concatena con las espiras de la bobina cambia y se produce una f.e.m. **e** proporcional a dicha velocidad. El signo de esta f.e.m. depende del sentido del movimiento y para lograr una dependencia aproximadamente lineal con la velocidad, los desplazamientos tienen que ser pequeños [12.2]. Esta limitación hace que este tipo de sensor se utilice fundamentalmente para medir vibraciones de frecuencias limitadas por su inercia.

12.3 Sensores electromagnéticos rotativos

Los sensores rotativos se emplean para medir **velocidad angular** o **posición angular** de un eje y poseen, al igual que las máquinas eléctricas (motores y generadores), un elemento giratorio (*rotor*) con un eje solidario al elemento al cual se desea medir su velocidad o posición angular y otro estático respecto al anterior (*estator*) coaxialmente ubicado respecto al rotor. El espacio que separa estos componentes metálicos, generalmente ocupado por aire, constituye un elemento de diferente reluctancia del resto de los materiales ferromagnéticos del *rotor* y el *estator*. Dicha reluctancia debe ser uniforme y lo más pequeña posible.

Los sensores electromagnéticos rotativos se diferencian de los generadores eléctricos en que el objetivo de los últimos es la conversión de energía mecánica a eléctrica con una elevada eficiencia, mientras que los

sensores electromagnéticos rotativos deben presentar, para el sistema mecánico al cual se acoplan, una carga pequeña para alterar lo menos posible las condiciones del proceso, por ello sus dimensiones son mucho menores y ofrecen una señal eléctrica de poca potencia.

Los sensores electromagnéticos de posición angular son transformadores rotativos en los que los devanados primarios y secundarios se encuentran en el *estator* y en el *rotor*. Uno de ellos (*rotor* o *estator*) actúa como primario y en el otro se induce una tensión de alterna de igual frecuencia que la excitación pero de amplitud dependiente del ángulo de desplazamiento relativo entre el *rotor* y el *estator*. Existe una dirección de alineación para la cual la tensión inducida en uno de los secundarios es cero, mientras que en el otro es máxima y respecto a ella se puede medir la posición angular del eje acoplado al rotor.

12.3.1 SENSORES ELECTROMAGNÉTICOS ROTATIVOS DE VELOCIDAD ANGULAR

Los sensores electromagnéticos de velocidad angular o tacómetros electromagnéticos son generadores especiales de corriente continua o de alterna. En función del tipo de señal eléctrica que generen (de continua o alterna) se denominan dinamos tacométricas y alternadores tacométricos, respectivamente. La figura 12.2 muestra el esquema básico de una dinamo tacométrica.

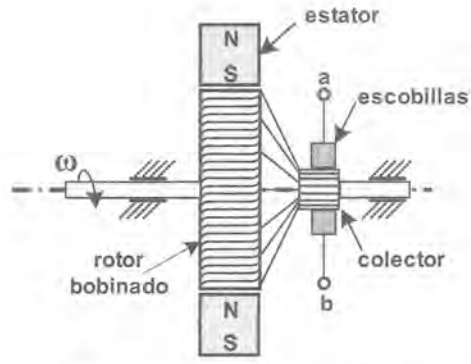


Figura 12.2. Dinamo tacométrica.

El estator está formado por un número de polos n, constituidos por imanes permanentes o por devanados por los que circula una corriente continua, de manera tal que el flujo magnético que producen es constante. Al girar el eje a una velocidad angular ω, las líneas de flujo magnético son "cortadas" por las espiras de la bobina del rotor induciéndose en ellas una tensión proporcional en magnitud a la velocidad angular y de polaridad dependiente del sentido de giro del eje. Las bobinas se conectan al colector giratorio y mediante unas escobillas fijas se conectan a los terminales de salida a y b. La forma de onda que se obtiene entre los terminales a y b se corresponde con la suma de un gran número de semi-senoides, por lo que contiene una componente de continua de gran valor y un rizado cuya frecuencia depende del número de polos y de la velocidad angular del rotor. Por esta razón su utilización para velocidades angulares pequeñas ofrece un gran error, aunque existe la posibilidad de acoplar su eje al del sistema a medir a través de un sistema de engranajes; no obstante, esto aumenta la carga sobre el sistema y no es aconsejable. Se comercializan con márgenes de salida de $0 \div 10V$ y $0 \div 5V$ y una gran gama de valores de la velocidad angular. Su utilización es cada vez menor debido a los problemas inherentes al mantenimiento y al desgaste de las escobillas y el colector.

Los **alternadores tacométricos** son generadores de corriente alterna que tienen la ventaja sobre los de continua de no utilizar escobillas ni colector, con lo que su mantenimiento es menos costoso. La figura 12.3 presenta el esquema simplificado de un alternador tacométrico de rotor con imanes permanentes[1].

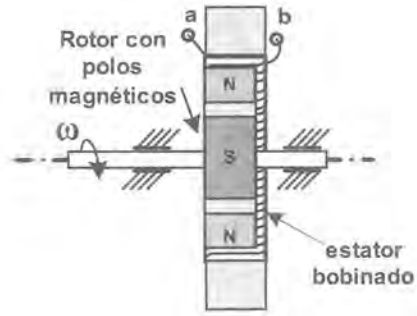


Figura 12.3. Alternador tacométrico de imán permanente en el rotor.

Al igual que en el tacómetro de continua, la tensión que se induce en el devanado del estator y que aparece entre los puntos a y b, es proporcional en magnitud y frecuencia a la velocidad angular del eje ω, pero es una señal alterna, por lo que si se quiere que la salida sea una señal de continua requiere de un sistema que realice la rectificación y el filtrado de esta señal senoidal. Los alternadores tacométricos comerciales actuales incorporan esta circuitería en el sensor y algunos ofrecen salidas de continua y de pulsos cuya frecuencia es proporcional a la velocidad angular por lo que, para la mayoría de las aplicaciones en las que la robustez, la fiabilidad y la resistencia a las vibraciones sean elementos determinantes, son preferidos en lugar de los de continua. La tabla 12.1 muestra las características generales de los tacómetros de la firma Vernitron's MAGTECHR [12.3].

PARÁMETRO	MAGNITUD
Diámetro exterior	25 ÷ 250 mm
Constante de tiempo	< 100 μs
Rizado de la salida	0,03%
Márgenes de entrada	1 ÷ 6000 rpm
Margen dinámico	1 ÷ 5000 mV/rpm

Tabla12.1. Características generales de los alternadores tacométricos de Vernitron's MAGTECHR.

Otro tipo de sensor electromagnético de velocidad angular de amplia utilización es el **tacogenerador de reluctancia variable**, presentado esquemáticamente en la figura 12.4.

Este sensor cuenta con una rueda dentada de material ferromagnético y de m dientes unida al eje cuya velocidad angular ω se desea medir. Esta rueda, al girar muy próxima a un devanado sobre un imán permanente, hace que el flujo Φ concatenado con la bobina varíe con el tiempo al variar la reluctancia $\Re$ del circuito magnético con la sucesión temporal de dientes-espacios de la rueda, de la forma que se presenta en la

[1] Este no es el único tipo de alternador tacométrico. También existe el de inducción en el que los polos del estator contienen dos devanados, uno excitado con una tensión alterna y otro en el que se obtiene la señal dependiente de la velocidad angular. El rotor está en cortocircuito, como en un motor de inducción. El lector puede obtener una información más completa al respecto en [12.1].

figura 12.4. Obsérvese que a reluctancia mínima (diente de la rueda enfrentado con el polo del imán) le corresponde flujo máximo en la bobina, y viceversa.

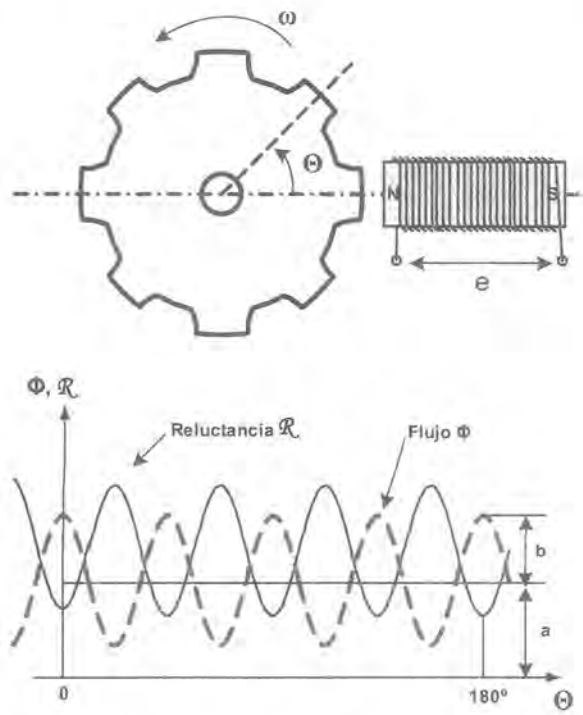


Figura 12.4. Tacogenerador de reluctancia variable.

El flujo Φ en función del ángulo de giro del disco puede aproximarse a una expresión de la siguiente forma [12.1]:

$$\Phi(\Theta) \approx a + b\cos\Theta \tag{12.2}$$

donde a es el valor medio del flujo Φ, b es la amplitud de la variación del flujo y Θ es el ángulo de giro del disco. Si suponemos que $\Theta = 0^\circ$ para $t = 0$, entonces se cumple en cada instante que $\Theta = \omega t$.

Si, con estas condiciones, se aplica la ley de Faraday, puede obtenerse:

$$e = b \cdot m \cdot \omega \cdot \text{sen}(m\omega t) \tag{12.3}$$

Puede observarse que tanto la amplitud como la frecuencia de esta tensión dependen directamente de la velocidad angular del disco, pero se prefiere utilizar la frecuencia debido a los problemas asociados al efecto de carga sobre la bobina si se utiliza como salida la tensión. Además, con un simple acondicionamiento puede convertirse esta señal en una onda cuadrada o en una sucesión de impulsos que pueden ser contados en un contador de pulsos remoto durante un tiempo fijo y obtener en ese tiempo un valor promedio de la velocidad angular del disco.

12.3.2 SENSORES ELECTROMAGNÉTICOS ROTATIVOS DE POSICIÓN ANGULAR

Los sensores electromagnéticos rotativos de posición angular más difundidos en la práctica por su gran precisión y robustez son los denominados "*resolvers*". Se trata de transformadores variables rotativos en los que

existen dos devanados en el estator (E_1 y E_2) dispuestos a 90° y otros dos en el rotor (R_1 y R_2) también a 90°, como se muestra esquemáticamente en la figura 12.5.

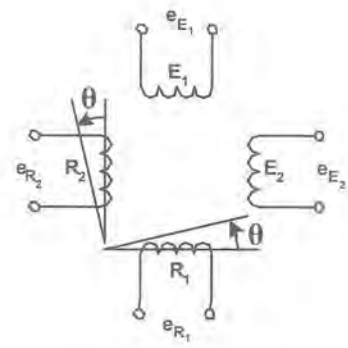


Figura 12.5. Esquema simplificado de los devanados de un *resolver*.

Si uno de los devanados del estator se excita con una señal de tensión senoidal de frecuencia f (por ejemplo, E_1) y el otro (E_2) no se alimenta o se pone en cortocircuito[2], por la disposición espacial relativa de los devanados existe una posición en la que la tensión senoidal inducida en el secundario del rotor R_1 es máxima, mientras que la tensión del secundario R_2 es nula. Si hacemos un desplazamiento del rotor hasta un ángulo θ, bajo estas condiciones, las tensiones inducidas en los secundarios serán:

$$e_{R_1} = e_{E_1} \cos\theta \tag{12.4}$$

$$e_{R_2} = e_{E_1} \operatorname{sen}\theta \tag{12.5}$$

De las expresiones anteriores podemos observar que el *resolver* ofrece tensiones en sus salidas cuya amplitud depende del ángulo de posición del eje del rotor respecto a la posición de referencia. La medida remota de posición angular también es posible con los *resolvers* denominados transmisores de datos (figura 12.6).

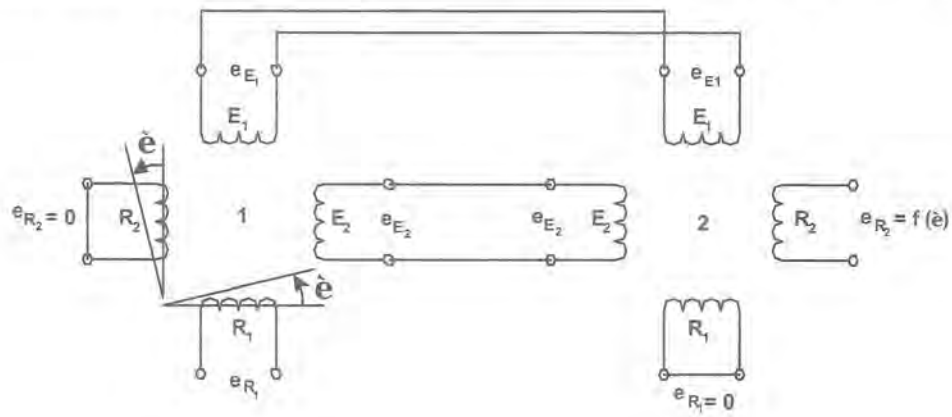


Figura 12.6. Aplicación de *resolvers* en la medida remota de posición angular.

[2] Al encontrarse este devanado espacialmente perpendicular al otro primario, la f.e.m. inducida en él teóricamente es nula, pero en la práctica se suele cortocircuitar para evitar posibles tensiones residuales inducidas.

Se interconectan adecuadamente los estatores mientras que la alimentación se realiza en uno de los devanados del rotor en el *resolver* 1 cuyo eje se encuentra solidario al objeto de medición de la posición angular è. La excitación puede ser la de red eléctrica u otra fuente senoidal en caso que se desee alimentar con una frecuencia mayor.

Si el eje del rotor del *resolver* se hace girar a una determinada velocidad angular ω_R y la excitación del primario es de una frecuencia ω, donde $\omega >> \omega_R$, entonces la señal senoidal de salida de frecuencia angular ω será modulada en amplitud con una señal senoidal de frecuencia angular ω_R. De esta forma, el *resolver* es también por excelencia un sensor de velocidad angular. Y, si a partir de una posición angular dada, se hace girar, en todo momento la señal de salida contiene la información de la posición angular del eje y la cantidad de vueltas que el mismo ha dado, lo que hace posible el incremento de la precisión del posicionamiento si se utilizan engranajes reductores, como es el caso del posicionamiento de antenas de radar.

A modo de ejemplo, la tabla 12.2 muestra algunas características de los *resolvers* fabricados por la firma Vernitron's MAGTECHR [12.3].

PARÁMETRO	MAGNITUD
Diámetro exterior	20 ÷ 300 mm
Frecuencia de excitación	60 ÷ 10000 Hz
Velocidad angular ω_R	Hasta 20000 r.p.m.
Tensión nula de salida	1 ÷ 3 mV/V
Exactitud	±10 arcseg ÷ ±4 arcmin

Tabla 12.2. Características de los *resolvers* de la firma Vernitron's MAGTECHR.

Como podemos observar de la tabla 12.2 los *resolvers* se comercializan en una gran variedad de tamaños y con un amplio margen de velocidades angulares del rotor. La relación de transformación suele ser unitaria, y la exactitud, en dependencia de la clase, puede satisfacer los más diversos requerimientos. Se caracterizan por su alta fiabilidad, una gran independencia de sus características con la temperatura, posibilidad de trabajo en ambientes no limpios con altas y muy bajas presiones. Sus aplicaciones más comunes se encuentran en robótica, en electrónica marina, aviónica e ingenios espaciales, donde resulta imposible el uso de otras técnicas de más bajo coste.

Resumen

Los sensores electromecánicos son sensores electromagnéticos en los que existe movimiento relativo entre los componentes del sensor. La salida de estos sensores se asocia a la velocidad lineal o a la angular de este movimiento relativo en dependencia del carácter del desplazamiento de las partes del sensor.

Los tacómetros electromagnéticos son generadores especiales de corriente continua o de alterna que ofrecen a su salida una señal eléctrica de tensión proporcional a la velocidad angular del eje del *rotor* y se caracterizan por ser sistemas robustos, fiables, resistentes a las vibraciones y con salidas muy independientes de la temperatura, presión y otros parámetros medioambientales, pero con muy poca utilización en la actualidad por su elevado coste inicial y de mantenimiento.

Los sensores electromagnéticos rotativos de posición angular más difundidos en la práctica por su gran precisión y robustez son los conocidos como "*resolvers*", que son transformadores variables rotativos que ofrecen tensiones en sus salidas cuya amplitud depende del ángulo de posición del eje del rotor respecto a una posición de referencia; al mismo tiempo este sensor es también por excelencia un sensor de velocidad angular.

13

Termopares

13.1 Principio de funcionamiento
13.2 Tipos de termopares
13.3 Curvas de calibración
13.4 Efectos de las uniones parásitas
13.5 Acondicionamiento de señal

13.1 Principio de funcionamiento

Un termopar es un sensor de temperatura constituido por dos metales diferentes cuya característica principal es que produce una tensión proporcional a la diferencia de temperaturas entre los puntos de unión de ambos metales. En el principio de funcionamiento del termopar están involucrados tres fenómenos:

- Efecto Seebeck.
- Efecto Peltier.
- Efecto Thomson.

El primero fue descubierto por Thomas Johan Seebeck en 1821, quien observó que si se tienen dos conductores diferentes formando un circuito cerrado y una de las uniones está a una temperatura T_1 y la otra a una temperatura diferente T_2, aparece una fuerza electromotriz que da lugar a la circulación de una corriente que se mantiene mientras las temperaturas sigan siendo diferentes. Si se abre el circuito, lo que se observa es la aparición de una tensión entre los terminales (figura 13.1).

Por otra parte, en 1834, Jean C. A. Peltier, descubrió un fenómeno inverso consistente en que, si se hace circular una corriente por el termopar, una de las uniones se calienta y la otra se enfría. Además, si se invierte el sentido de la circulación de la corriente, la unión que se calienta y la que se enfría se invierten.

Además, William Thomson (Lord Kelvin) descubrió en 1854 que cuando se hace circular una corriente constante a través de un conductor en el que uno de sus extremos se mantiene a diferente temperatura que el otro, se produce una transferencia de calor proporcional, aproximadamente, al producto de la corriente por el gradiente de temperatura. Si se invierte el sentido de la circulación de la corriente, también se invertirá el sentido de la transferencia de calor. Observe que este efecto es diferente al efecto Joule en el que la transferencia de calor es proporcional al cuadrado de la corriente (I^2R) y no admite la inversión.

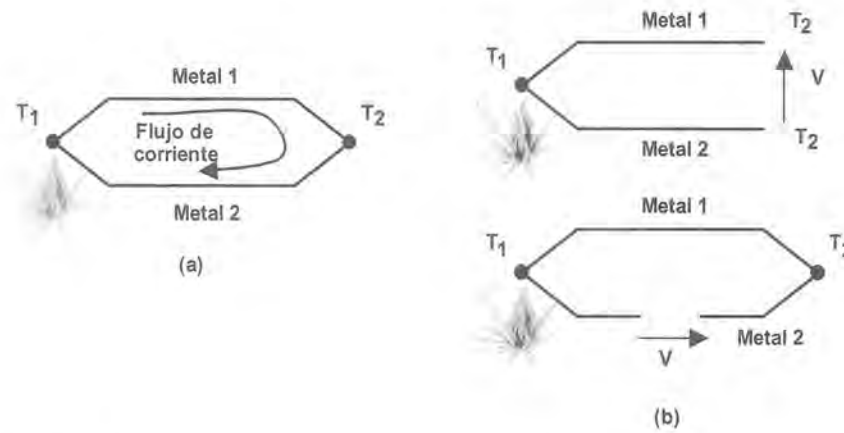


Figura 13.1. Efecto Seebeck. (a) Al calentar una de las uniones con respecto a la otra se observa la circulación de una corriente proporcional a la diferencia de temperaturas; (b) si se abre el circuito, se observa la aparición de una tensión entre los terminales.

Del estudio experimental de los termopares se dedujeron tres leyes, denominadas leyes termoeléctricas, que resumen su comportamiento [13.6]:

- Ley de los circuitos homogéneos: la tensión generada por un termopar cuyas uniones se encuentran a las temperaturas T_1 y T_2 no depende de la temperatura a la que se encuentren los puntos intermedios (figura 13.2).

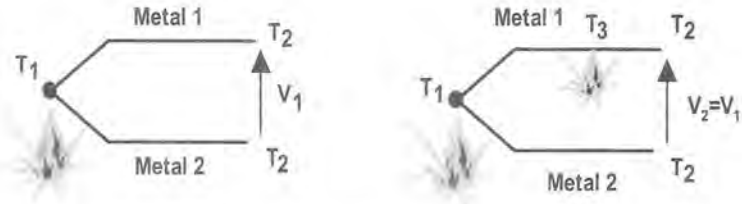


Figura 13.2. Ley de los circuitos homogéneos.

- Ley de los metales intermedios: si se introduce un tercer metal en serie con uno de los que constituyen el termopar, la tensión generada por el termopar no varía siempre que los extremos del metal insertado se encuentren a la misma temperatura (figura 13.3).

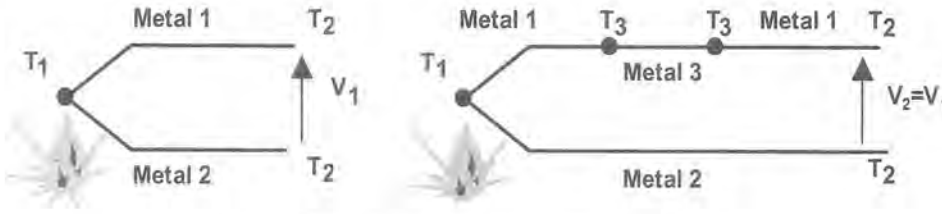


Figura 13.3. Ley de los metales intermedios.

- Ley de las temperaturas intermedias: si $V_{T1,T2}$ es la tensión generada por un termopar cuyas uniones están a las temperaturas T_1 y T_2, y $V_{T2,T3}$ es la tensión cuando están a T_2 y T_3, la f.e.m., $V_{T1,T3}$, cuando están a las temperaturas T_1 y T_3 es igual a $V_{T1,T2} + V_{T2,T3}$ (figura 13.4).

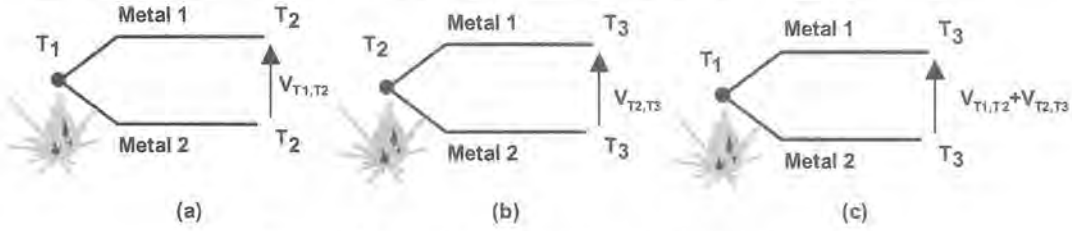


Figura 13.4. Ley de las temperaturas intermedias.

No obstante, el efecto termoeléctrico puede ser explicado de forma teórica a partir de la estructura atómica del material [13.4]. En general, los electrones situados en el nivel más externo están unidos débilmente al núcleo. Cuando se calienta un conductor en un extremo, estos electrones aumentan su energía y tienden a alcanzar el otro extremo por mecanismos de difusión, de forma que el extremo frío adquiere carga negativa y el extremo caliente, positiva. Esto provoca, a su vez, un campo eléctrico que tiende a oponerse a la difusión de forma que se llega a un estado de equilibrio.

Por tanto, el efecto termoeléctrico no se debe a la existencia de dos metales diferentes sino a la diferencia de temperaturas entre dos puntos de un mismo metal. No obstante, la fuerza electromotriz generada no se podría medir si se tuviera un único metal. Así, si se calienta un punto de único metal (punto A en la figura 13.5) la tensión entre los terminales extremos B y C es nula, puesto que la tensión V_{AB} es igual a la caída V_{AC} dado que los extremos de ambas ramas están a la misma temperatura.

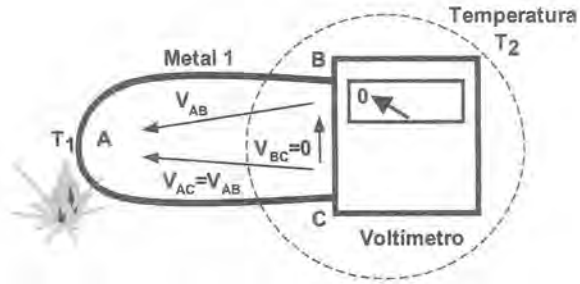


Figura 13.5. La tensión generada por el efecto termoeléctrico entre los extremos de un único metal en el que se caliente un punto intermedio es nula.

Por ello, la medición ha de realizarse sobre circuitos asimétricos, es decir, constituidos con dos o más metales diferentes. Como la caída de tensión depende del material (de su estructura electrónica) además de la diferencia de temperaturas, si se dispone de dos metales distintos la fuerza electromotriz en cada uno de ellos será diferente y la fuerza electromotriz resultante será distinta de cero.

13.2 Tipos de termopares

Aunque el termopar puede construirse con dos metales diferentes cualesquiera, la necesidad de una elevada sensibilidad, estabilidad a lo largo del tiempo, linealidad, etc. ha llevado a que tan sólo se utilicen determi-

nados materiales dando lugar a diferentes tipos. Algunos de los más importantes son: J, K, N, T, R, S y B (tabla 13.1), aunque existen otros como, por ejemplo, el E y el W. Los cuatro primeros se conocen con el nombre de termopares de metales base porque están hechos de metales comunes como el cobre, níquel, aluminio, etc. mientras que los termopares R, S y B se conocen como termopares de metales nobles porque están hechos de platino y rodio. Estos últimos se utilizan cuando las temperaturas a medir son muy elevadas y exceden los márgenes que se pueden alcanzar con los metales base o cuando hay problemas de compatibilidad química (tabla 13.2).

Tipo	Composición (terminal positivo - negativo)	Campo de medida recomendado	Sensibilidad (a 25°C)
J	Fe - Constantán*	0 a 760°C	51,5 μV/°C
K	Cromel* - Alumel*	–200 a 1250°C	40,5 μV/°C
N	Nicrosil* - Nisil*	0 a 1260°C	26,5 μV/°C
T	Cu - Constantán	–200 a 350°C	41,0 μV/°C
R	13%Pt 87%Rh - Pt	0 a 1450°C	6 μV/°C
S	10%Pt 90%Rh - Pt	0 a 1450°C	6 μV/°C
B	30%Pt 70%Rh - 6%Pt 94%Rh	800 a 1800°C	9 μV/°C (a 1000 °C)

Tabla 13.1. Tipos de termopares y características fundamentales. (Elaborada según [13.2], [13.7])
*Constantán: 55%Cu 45%Ni, Cromel: 90%Ni 10%Cr, Alumel: 95%Al 5%Ni,
Nicrosil: 84,6%Ni 14,2%Cr 1,4%Si, Nisil: 95,5%Ni 4,4%Si 1,0%Mg

Tipo	Aplicabilidad
J	Apropiado para atmósferas inertes o reductoras. Las atmósferas oxidantes disminuyen la vida útil debido a la presencia de hierro en el termopar que, además, se oxida muy rápidamente por encima de 538°C. No es adecuado para bajas temperaturas (por debajo de 0°C).
K	Muy utilizado por encima de 538°C debido a las limitaciones del termopar de tipo J. El cromo tiende a oxidarse ante la presencia de oxígeno lo que puede llevar a importantes derivas en el margen de 816 a 1038°C.
N	Se utiliza en aplicaciones donde el termopar de tipo K tiene problemas de oxidación.
T	Adecuado para atmósferas oxidantes, inertes y reductoras.
R, S	Recomendados para altas temperaturas. El de tipo R se utiliza industrialmente, mientras que el S se usa en laboratorios. El uso continuado a altas temperaturas provoca el crecimiento del granulado y puede sufrir una ruptura mecánica. Deben protegerse con tubos no metálicos y aislantes cerámicos. Tienden a descalibrarse debido a la difusión del rodio a la rama de platino puro y a su volatilización.
B	Semejante a los tipos R y S aunque el límite de temperatura es mayor y es menos susceptible al crecimiento del granulado.

Tabla 13.2. Aplicabilidad de los diferentes tipos de termopares (tomado de [13.2]).

La unión entre las dos ramas que componen un termopar, también denominada unión caliente, puede estar realizada mediante el simple trenzado de los cables, fusión o incluso mediante la soldadura. Aunque para soldar es necesario emplear un metal adicional, la tensión observada no varía como se puede deducir aplicando la ley de los metales intermedios.

Por otra parte, se distinguen tres tipos de uniones según el encapsulado de las mismas [13.6] (figura 13.6):

- Unión a masa: la unión hace contacto con la cápsula. Suele ser la más utilizada debido a que el termopar queda protegido del ambiente sin que se reduzca excesivamente el tiempo de respuesta (unos 2 s).
- Unión aislada: la unión está separada de la cápsula mediante aislante eléctrico de elevada conductividad térmica (normalmente óxido de magnesio). Resulta adecuado, por ejemplo, para medir temperaturas de líquidos conductores. El tiempo de respuesta suele rondar los 5 s.
- Unión desnuda o expuesta: la unión no está encapsulada. Presenta la ventaja de que el tiempo de respuesta puede ser muy reducido (0,1 s) pero el termopar se degrada rápidamente cuando las temperaturas son elevadas.

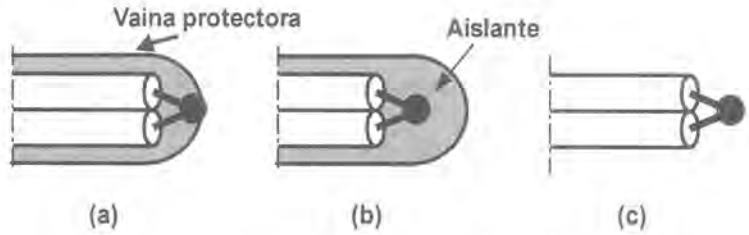


Figura 13.6. Tipos de uniones. (a) A masa; (b) aislada; (c) expuesta.

13.3 Curvas de calibración

La tensión entre los terminales del termopar depende de la temperatura de las dos uniones, es decir, de la temperatura de la unión situada en la zona que se desea medir (denominada unión caliente[1]) y de la propia temperatura a la que se encuentra la unión con el equipo de medición (unión fría); por lo tanto, para cada temperatura de la unión fría existe una curva de calibración (relación de la tensión con la temperatura objeto de la medida). No obstante, conociendo la curva de calibración para una temperatura determinada de una de las uniones, se puede conocer la curva de calibración para cualquier temperatura aplicando la ley de las temperaturas intermedias (véase el ejercicio 13.1). Así, las curvas de calibración de los termopares se encuentran recogidas en una serie de tablas [13.1] que muestran la tensión del termopar suponiendo que una de las uniones está a 0°C.

EJERCICIO 13.1

Se dispone de un termopar que presenta la curva de calibración mostrada en la tabla 13.3 cuando una de las uniones se mantiene a 0°C.

a) Si la temperatura de una de las uniones es de 145°C y la unión con el equipo de medición es de 20°C, ¿cuál es la tensión medida?

b) Si la tensión medida fuera de 3,422 mV, ¿a qué temperatura está la unión suponiendo que el equipo de medición sigue a 20°C?

[1] Aunque la unión objeto de la medida se suele denominar como unión caliente y la correspondiente al equipo de medición, unión fría, esto no implica necesariamente que aquella sea siempre superior a esta última.

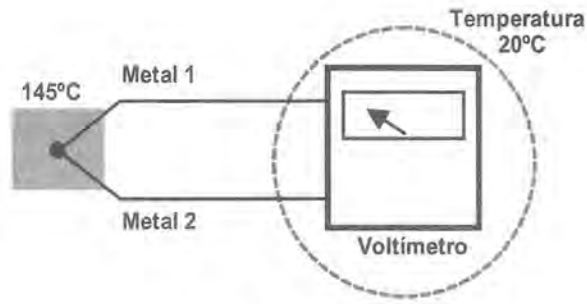


Figura 13.7

ºC	0	1	2	3	4	5	6	7	8	9	10	ºC
						Tensión termoeléctrica (mV)						
0	0.000	0.039	0.079	0.119	0.158	0.198	0.238	0.277	0.317	0.357	0.397	0
10	0.397	0.437	0.477	0.517	0.557	0.597	0.637	0.677	0.718	0.758	0.798	10
20	0.798	0.838	0.879	0.919	0.960	1.000	1.041	1.081	1.122	1.163	1.203	20
30	1.203	1.244	1.285	1.326	1.366	1.407	1.448	1.489	1.530	1.571	1.612	30
40	1.612	1.653	1.694	1.735	1.776	1.817	1.858	1.899	1.941	1.982	2.023	40
50	2.023	2.064	2.106	2.147	2.188	2.230	2.271	2.312	2.354	2.395	2.436	50
60	2.436	2.478	2.519	2.561	2.602	2.644	2.685	2.727	2.768	2.810	2.851	60
70	2.851	2.893	2.934	2.976	3.017	3.059	3.100	3.142	3.184	3.225	3.267	70
80	3.267	3.308	3.350	3.391	3.433	3.474	3.516	3.557	3.599	3.640	3.682	80
90	3.682	3.723	3.765	3.806	3.848	3.889	3.931	3.972	4.013	4.055	4.096	90
100	4.096	4.138	4.179	4.220	4.262	4.303	4.344	4.385	4.427	4.468	4.509	100
110	4.509	4.550	4.591	4.633	4.674	4.715	4.756	4.797	4.838	4.879	4.920	110
120	4.920	4.961	5.002	5.043	5.084	5.124	5.165	5.206	5.247	5.288	5.328	120
130	5.328	5.369	5.410	5.450	5.491	5.532	5.572	5.613	5.653	5.694	5.735	130
140	5.735	5.775	5.815	5.856	5.896	5.937	5.977	6.017	6.058	6.098	6.138	140
150	6.138	6.179	6.219	6.259	6.299	6.339	6.380	6.420	6.460	6.500	6.540	150
160	6.540	6.580	6.620	6.660	6.701	6.741	6.781	6.821	6.861	6.901	6.941	160
170	6.941	6.981	7.021	7.060	7.100	7.140	7.180	7.220	7.260	7.300	7.340	170
180	7.340	7.380	7.420	7.460	7.500	7.540	7.579	7.619	7.659	7.699	7.739	180
190	7.739	7.779	7.819	7.859	7.899	7.939	7.979	8.019	8.059	8.099	8.138	190
200	8.138	8.178	8.218	8.258	8.298	8.338	8.378	8.418	8.458	8.499	8.539	200

Tabla 13.3. Tensión entre los terminales de un termopar de tipo K en función de la temperatura de la unión cuando la unión de referencia está a 0ºC (extraída de [13.5]).

SOLUCIÓN:

Aplicando la ley de las temperaturas intermedias:

$$V_{145,0} + V_{0,20} = V_{145,20}$$

$$V_{145,0} - V_{20,0} = V_{145,20}$$

De la tabla obtiene que $V_{145,0} = 5{,}937$ mV y $V_{20,0} = 0{,}798$ mV. Por la tanto, la tensión medida será de 5,139 mV.

Por otra parte, si la tensión medida $V_{x,20}$, cuando una unión está a una temperatura x, y la otra a 20°C, es de 3,422 mV:

$$V_{x,0} - V_{20,0} = V_{x,20}$$

$$V_{x,0} = V_{x,20} + V_{20,0}$$

$$V_{x,0} = 3{,}422 + 0{,}798 = 4{,}220 \text{ mV}$$

De la tabla 13.3 se obtiene que esta tensión corresponde a una temperatura de la zona caliente igual a 103°C.

Las tablas de calibración de los termopares vienen expresadas en relación con aproximaciones polinómicas (tabla 13.4). Además, las propias normas que recogen las tablas de calibración, recogen las funciones inversas que permiten determinar, en función de la tensión observada, la temperatura de la unión caliente *con un pequeño error (tabla 13.5).*

Margen de aplicación	–210 ° C a 76°C	760°C a 1200°C
Tensión en el termopar (T: temperatura de la unión caliente en °C; unión fría a 0°C)	$V = \sum_{i=1}^{n} a_i \cdot T^i$ μV $a_1 = 5{,}0381187815 \cdot 10^{1}$ $a_2 = 3{,}0475836930 \cdot 10^{-2}$ $a_3 = -8{,}5681065720 \cdot 10^{-5}$ $a_4 = 1{,}3228195295 \cdot 10^{-7}$ $a_5 = -1{,}7052958337 \cdot 10^{-10}$ $a_6 = 2{,}0948090697 \cdot 10^{-13}$ $a_7 = -1{,}2538395336 \cdot 10^{-16}$ $a_8 = 1{,}5631725697 \cdot 10^{-20}$	$V = \sum_{i=0}^{n} a_i \cdot T^i$ μV $a_0 = 2{,}9645625681 \cdot 10^{5}$ $a_1 = -1{,}4976127786 \cdot 10^{3}$ $a_2 = 3{,}1787103924$ $a_3 = -3{,}1847686701 \cdot 10^{-3}$ $a_4 = 1{,}5720819004 \cdot 10^{-6}$ $a_5 = -3{,}0691369056 \cdot 10^{-10}$

Tabla 13.4. Funciones polinómicas que describen la curva de calibración de un termopar de tipo J (tomado de [13.1]).

Margen de aplicación	–8 095 μV a 0 μV (–210 ° C a 0°C)	0 μV a 42 919 μV (0°C a 760°C)	42919 μV a 69553 μV (760°C a 1200°C)
Error (°C)	Máx. = 0, 028 Mín. = –0, 048	Máx. = 0, 037 Mín. = –0, 035	Máx. = 0, 024 Mín. = –0, 037
Temperatura de la unión caliente (V: tensión del termopar en μV; unión fría a 0°C)	$T = \sum_{i=1}^{n} d_i \cdot V^i$ °C $d_1 = 1{,}9528268 \cdot 10^{-2}$ $d_2 = -1{,}2286185 \cdot 10^{-6}$ $d_3 = -1{,}0752178 \cdot 10^{-9}$ $d_4 = -5{,}9086933 \cdot 10^{-13}$ $d_5 = -1{,}7256713 \cdot 10^{-16}$ $d_6 = -2{,}8131513 \cdot 10^{-20}$ $d_7 = -2{,}3963370 \cdot 10^{-24}$ $d_8 = -8{,}3823321 \cdot 10^{-29}$	$T = \sum_{i=1}^{n} d_i \cdot V^i$ °C $d_1 = 1{,}978425 \cdot 10^{-2}$ $d_2 = -2{,}001204 \cdot 10^{-7}$ $d_3 = 1{,}036969 \cdot 10^{-11}$ $d_4 = -2{,}549687 \cdot 10^{-16}$ $d_5 = 3{,}585153 \cdot 10^{-21}$ $d_6 = -5{,}344285 \cdot 10^{-26}$ $d_7 = 5{,}099890 \cdot 10^{-31}$	$T = \sum_{i=0}^{n} d_i \cdot V^i$ °C $d_0 = -3{,}11358187 \cdot 10^{3}$ $d_1 = 3{,}00543684 \cdot 10^{-1}$ $d_2 = -9{,}94773230 \cdot 10^{-6}$ $d_3 = 1{,}70276630 \cdot 10^{-10}$ $d_4 = -1{,}43033468 \cdot 10^{-15}$ $d_5 = 4.73886084 \cdot 10^{-21}$

Tabla 13.5. Funciones polinómicas inversas que permiten determinar la temperatura en función de la tensión medida para un termopar de tipo J (tomado de [13.1]).

En general, las curvas de calibración son bastante lineales (figura 13.8), especialmente las correspondientes a los metales base. Por ello, en muchos casos la repuesta se aproxima a una línea recta con un error aceptablemente pequeño (dependiendo, fundamentalmente, del campo de medida).

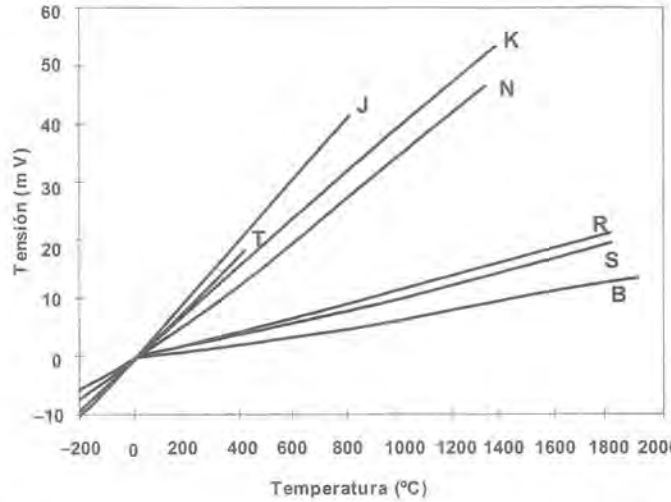


Figura 13.8. Tensión termoeléctrica en los diferentes termopares en función de la temperatura de una unión cuando la otra se mantiene a 0°C.

EJERCICIO 13.2

Proponga una linealización para la curva de calibración de un termopar de tipo K en el margen 0 - 100°C y determine el error de linealidad cometido.

SOLUCIÓN:

La linealización se va a realizar mediante una recta que pase por los extremos del margen de medida (0 y 100°C). La sensibilidad, S, según estos puntos extremos es (tabla 13.3):

$$S = \frac{4{,}096\ \text{mV} - 0\ \text{mV}}{100^{\circ}\text{C}} = 40{,}96\ \mu\text{V}/^{\circ}\text{C}$$

Teniendo en cuenta que para 0°C la tensión en el termopar es de 0 mV, la línea recta que se propone es:

$$V = S \cdot T_c$$

donde T_c es la temperatura de la unión caliente y V la tensión proporcionada por el termopar.

La máxima desviación de la curva linealizada con respecto a la curva real tiene lugar para una temperatura de 34°C. La tabla de calibración indica que la tensión a esta temperatura es de 1,366 mV, mientras que la respuesta linealizada ofrece un valor de 1,392 mV. Por tanto, el error en % sobre el fondo de escala de salida es:

$$\text{error} = \frac{1{,}392\ \text{mV} - 1{,}366\ \text{mV}}{4{,}096\ \text{mV}} \cdot 100 = 0{,}65\%$$

lo que corresponde a un error máximo de 0,65°C en el campo de medida especificado.

13.4 Efectos de las uniones parásitas

Para medir la tensión en un termopar se necesita unir los terminales del mismo con el equipo de medida cuyas conexiones estarán además realizadas, en general, de diferentes materiales a los que constituyen el termopar (por ejemplo, de cobre) (figura 13.9). Se puede demostrar muy fácilmente a partir de la ley de los metales intermedios que, si las nuevas uniones están a la misma temperatura, la tensión medida en estas condiciones corresponde a la que presenta el termopar original pero siendo la unión fría el punto de contacto con los terminales de contacto. La tensión termoeléctrica entre cada una de las nuevas uniones y el equipo es idéntica por lo que se cancelan en la medida. Por el mismo motivo, las soldaduras internas del polímetro no influyen si se encuentran a la misma temperatura. En definitiva, y como ya se ha mencionado, los circuitos simétricos no dan lugar a variaciones en la medida.

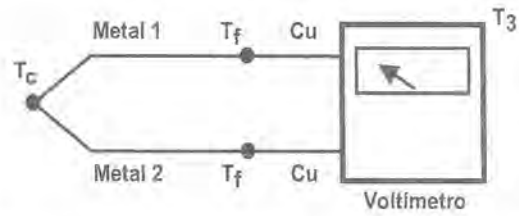


Figura 13.9. Conexión de un termopar a un voltímetro. La tensión medida depende únicamente de las temperaturas de la uniones caliente y fría, T_c y T_f, respectivamente.

Para garantizar que las uniones con el equipo se encuentren a la misma temperatura se suelen utilizar unos conectores especiales isotérmicos que garantizan una distribución uniforme de la temperatura de forma que esta sea idéntica en las dos uniones (figura 13.10).

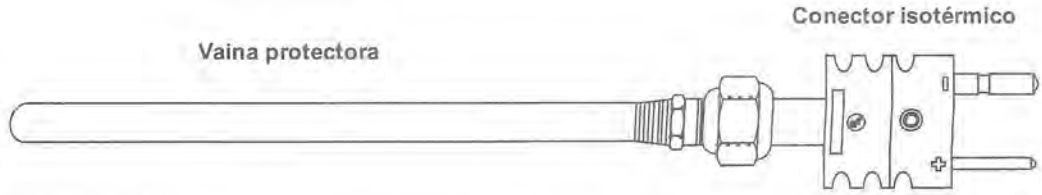


Figura 13.10. Aspecto de un termopar comercial con la vaina protectora y el conector isotérmico.

13.5 Acondicionamiento de señal

El termopar proporciona una tensión de pequeño valor proporcional a la diferencia de temperaturas entre dos uniones. Además, la sensibilidad es del orden de algunos microvoltios por grado centígrado. Así pues, en el acondicionamiento de señal habrá que tener presentes dos aspectos básicos:

1. El circuito de acondicionamiento consistirá, básicamente, en un amplificador de tensión en el que serán importantes aspectos tales como la tensión de desviación y sus derivas.
2. Para determinar la temperatura objeto de la medida será necesario conocer la temperatura de la otra unión.

Para resolver esta última cuestión existen tres posibilidades:

- Compensación mediante uniones de referencia.
- Compensación digital.
- Compensación analógica.

Compensación mediante uniones de referencia

Consiste en fijar en un valor conocido la temperatura de una de las uniones. Tradicionalmente, se introduce en un baño de hielo y agua (figura 13.11) para garantizar 0°C. Bien construido, el error puede ser de unos 0,001°C [13.3]. Se trata de un método de referencia pero es difícil de llevar a cabo por lo que se suele utilizar únicamente en laboratorios para realizar calibraciones.

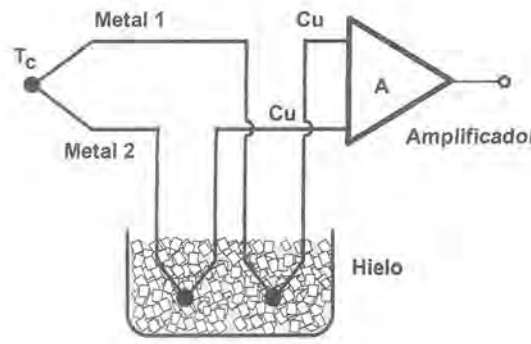


Figura 13.11. Compensación de la unión fría mediante un baño de hielo.

Compensación digital

Consiste en medir la temperatura de la unión fría mediante otro sensor de temperatura (por ejemplo, una Pt100 o una NTC) (figura 13.12).

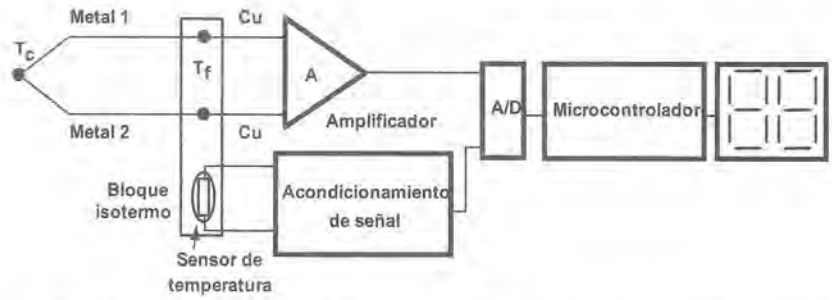


Figura 13.12. Compensación digital o compensación software de la unión fría.

Una vez conocida la temperatura de la unión fría y la tensión que proporciona el termopar se puede determinar fácilmente la temperatura de la unión caliente (véase, por ejemplo, el ejercicio 13.1). Además, la propia utilización del microcontrolador permite utilizar aproximaciones polinómicas a la curva de calibración consiguiéndose una elevada exactitud.

Compensación analógica

Las curvas de calibración de los termopares habituales muestran la tensión que proporcionan siempre que la unión fría esté a 0°C. Para una temperatura de la unión caliente T_c, y de la unión fría T_f, cualquiera, la tensión proporcionada V_{T_c,T_f} sería, aplicando la ley de las temperaturas intermedias:

$$V_{T_c,T_f} = V_{T_c,0} + V_{0,T_f} \tag{13.1}$$

por lo tanto:

$$V_{T_c,0} = V_{T_c,T_f} + V_{T_f,0} \tag{13.2}$$

Así pues el objetivo de la compensación analógica (o compensación hardware) es sumar a la tensión que proporciona el termopar V_{T_c,T_f}, la tensión $V_{T_f,0}$ que correspondería al mismo termopar en el que las temperaturas de las uniones fueran T_f y 0°C. A partir de la tensión resultante se puede determinar la temperatura de la zona caliente aplicando directamente la curva de calibración. Para ello será necesario, evidentemente, medir la temperatura de la zona fría con otro sensor y realizar un circuito de acondicionamiento que proporcione la tensión a sumar (figura 13.13).

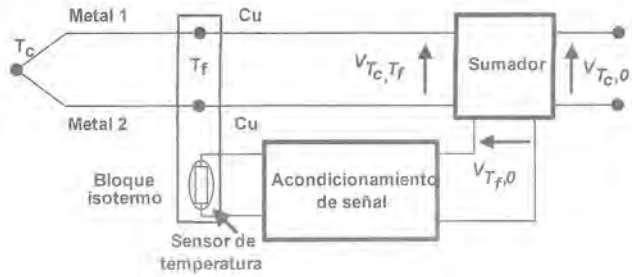


Figura 13.13. Compensación analógica de la unión fría.

EJERCICIO 13.3

Diseñe un circuito de acondicionamiento para un termopar de tipo K para un campo de medida entre 0-100°C de forma que su salida esté comprendida en el margen 0-1 V. Se compensará la unión fría analógicamente mediante una Pt100 (suponga que la unión fría puede estar comprendida entre 10 y 30°C).

SOLUCIÓN:

Linealizando la respuesta del termopar en el margen 0-100°C al igual que en el ejercicio 13.2, la tensión proporcionada por el termopar cuando la unión fría está a 0°C es:

$$V_{T_c,0} = 40,96 \cdot T_c \ \mu V$$

donde T_c es la temperatura de la unión caliente.

Por otra parte, la tensión, $V_{T_f,0}$, se puede linealizar en el margen 10-30°C por la expresión:

$$V_{T_f,0} = 40,00 \cdot T_f \ \mu V$$

Esta tensión se obtendrá, por ejemplo, a partir de una Pt100 mediante un circuito en puente como el de la figura 13.14.

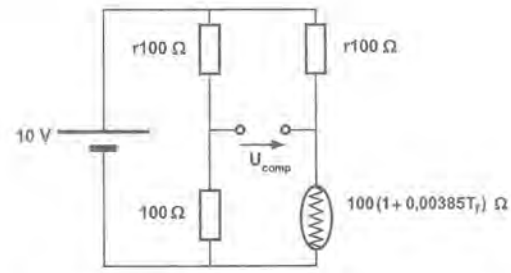


Figura 13.14. Circuito para medir la temperatura de la unión fría.

Diseñando el circuito mediante las técnicas habituales se obtiene que, tomando r = 960, resulta que la tensión $V_{comp} = V_{T_f,0}$. La suma de esta tensión y la del termopar puede realizarse con la configuración de la figura 13.15a y el circuito se completaría con un amplificador de instrumentación para amplificar la tensión al margen de salida pedido y con un bloque isotermo para realizar las conexiones (figura 13.15b).

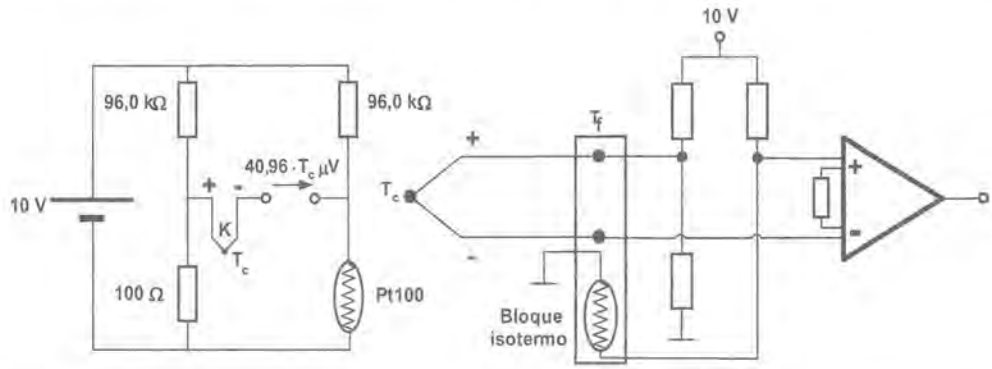


Figura 13.15. (a) Circuito para el acondicionamiento de señal del termopar que proporciona una tensión dependiente únicamente de la temperatura de la unión caliente, T_c; (b) amplificación de la tensión para la adaptación al margen de salida pretendido y conexionado del termopar y de la Pt100.

Resumen

Un termopar es un sensor de temperatura constituido por dos metales diferentes cuya característica principal es que produce una tensión proporcional a la diferencia de temperaturas entre los puntos de unión de ambos metales.

El funcionamiento de los termopares se caracteriza a partir de tres leyes: la ley de los circuitos homogéneos, la ley de los metales intermedios y la ley de las temperaturas intermedias.

Se distinguen varios tipos de termopares en función de los metales que los componen. Entre los más comunes están los de tipo J y los de tipo K. También existen algunos termopares hechos con metales nobles y que son adecuados para atmósferas agresivas.

El comportamiento de los diferentes tipos de termopares está perfectamente caracterizado en las normas.

En el acondicionamiento de señal hay que tener en cuenta que para determinar la temperatura objeto de la medida se necesita fijar o conocer la temperatura a la que se encuentra la unión fría del termopar. Para esto último se suele emplear otro sensor de temperatura adicional.

Problemas propuestos

13.1 Se dispone de un termopar de tipo J que se pretende conectar a un circuito de medida que ya incorpora una NTC para compensar la unión fría. Sin embargo, los cables del termopar son excesivamente cortos como para alcanzar el punto de medida pretendido. ¿Cómo se deberían prolongar los cables?

13.2 La resistencia de salida de un termopar, ¿es grande o pequeña?

13.3 Si se conectan varios termopares idénticos en serie de forma que las uniones calientes de todos ellos estén situadas en una misma zona y las uniones frías también estén situadas en una misma zona, aunque diferente de la anterior, ¿cuál es la tensión que se mide en los extremos del conjunto?

13.4 Diseñe un circuito de acondicionamiento para termopares de tipo K, empleando una NTC para compensar la unión fría. (Resistencia de la NTC: 100 kΩ a 25°C, B = 4190K)

14

Sensores piezoeléctricos

14.1 Introducción
14.2 Comportamiento de los materiales piezoeléctricos
14.3 Los dispositivos piezoeléctricos como sensores
14.4 Sensores de ultrasonidos

14.1 Introducción

El efecto piezoeléctrico[1] fue descubierto por Jacques y Pierre Curie en 1880 en los cristales de cuarzo. Este fenómeno se produce en determinados sistemas cristalinos y consiste en la aparición de una cierta carga eléctrica cuando el cristal es sometido a algún tipo de deformación.

Además del cristal de cuarzo, existen otras sustancias naturales que poseen comportamientos similares tales como la turmalina o la sal de Rochelle. Estas sustancias tienen una gran estabilidad tanto ante el cambio en las condiciones medioambientales como en el transcurso del tiempo, pero las señales que son capaces de producir resultan muy débiles lo que es un problema para cualquier sistema de instrumentación.

Habida cuenta de la importancia del fenómeno y de su aplicabilidad (ya en 1920 Langevin desarrolló un sistema de emisión/recepción de señales acústicas en agua que fue, de hecho, el primer sistema SONAR) se empezó a trabajar en el desarrollo de cristales que exhibieran las mismas propiedades que estos materiales naturales pero que mejoraran su comportamiento desde el punto de visa de la potencia de las señales producidas. Así, se obtuvieron materiales como el titanato de bario o el titanato-circonato de plomo (PZT) que, si bien no tienen tanta estabilidad como las sustancias naturales, sí que presentan un nivel mucho más alto de señal de salida lo que redunda directamente en una mejora de los parámetros más significativos de cualquier sistema de instrumentación.

[1] El término "piezoeléctrico" tiene origen griego: "piezo" significa presión, por lo que se podría asignar un significado etimológico de electricidad producida por la presión.

Los materiales sintéticos producen el efecto piezoeléctrico gracias a la anisotropía de su estructura (figura 14.1) y a la distribución de sus cargas eléctricas [14.8] que se suele conseguir después de un tratamiento [14.14]: los dominios[2] pueden estar orientados en cualquier dirección de tal manera que, en condiciones generales, el cristal tendrá un comportamiento macroscópico nulo desde el punto de vista eléctrico y de deformación. Sin embargo, después de un proceso de polarización con campos eléctricos muy intensos, se puede conseguir una reordenación de los dominios y el reforzamiento de los comportamientos anisótropos que permite obtener el efecto piezoeléctrico (figura 14.2).

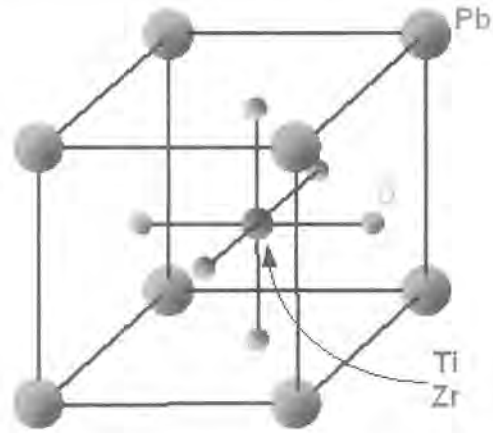


Figura 14.1. Estructura cristalina de un material sintético, el PZT.

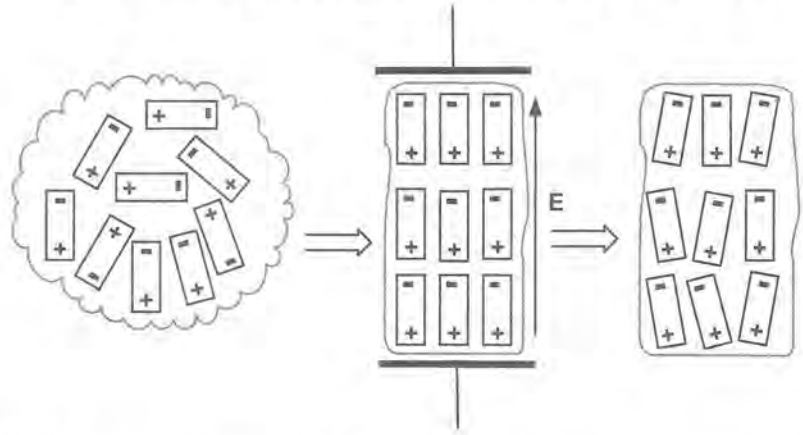


Figura 14.2. Obtención de un material piezoeléctrico por reordenación de los dominios magnéticos bajo una fuerte polarización.

Este comportamiento se manifiesta por debajo de la temperatura de Curie del material ya que si se alcanza este valor, se produce un cambio en la estructura cristalina que pasa a ser isotrópica con lo que desaparece el fenómeno. Pero el paso por la temperatura de Curie no sólo provoca una desaparición momentánea de este efecto, sino que produce cambios permanentes ya que al descender hasta menores temperaturas, se produce una re-cristalización en las mismas condiciones originales pero con dominios orientados de forma aleatoria con lo que el cristal pierde su comportamiento macroscópico (figura 14.3).

[2] Un dominio es un grupo de unidades cristalográficas con la misma polarización y la misma orientación de deformación.

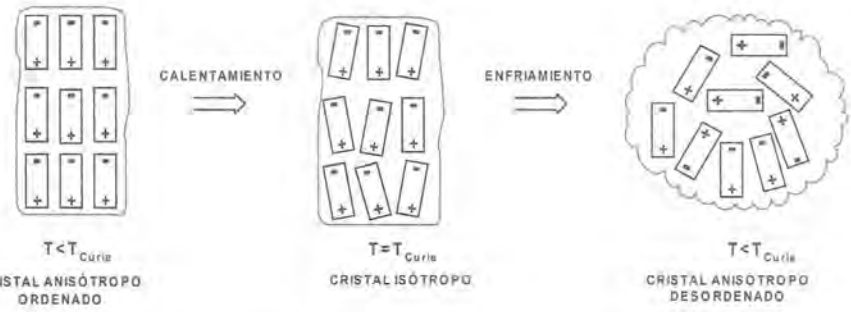


Figura 14.3. Efecto del calentamiento hasta la temperatura de Curie sobre la distribución de los dominios magnéticos en un cristal anisótropo ordenado.

*El proceso de fabricación de los materiales cerámicos utilizados como dispositivos piezoeléctricos supo*ne una serie de pasos que se indican en la figura 14.4, con etapas de sinterizado, pulido y polarizado para concluir con la deposición de los electrodos [14.8].

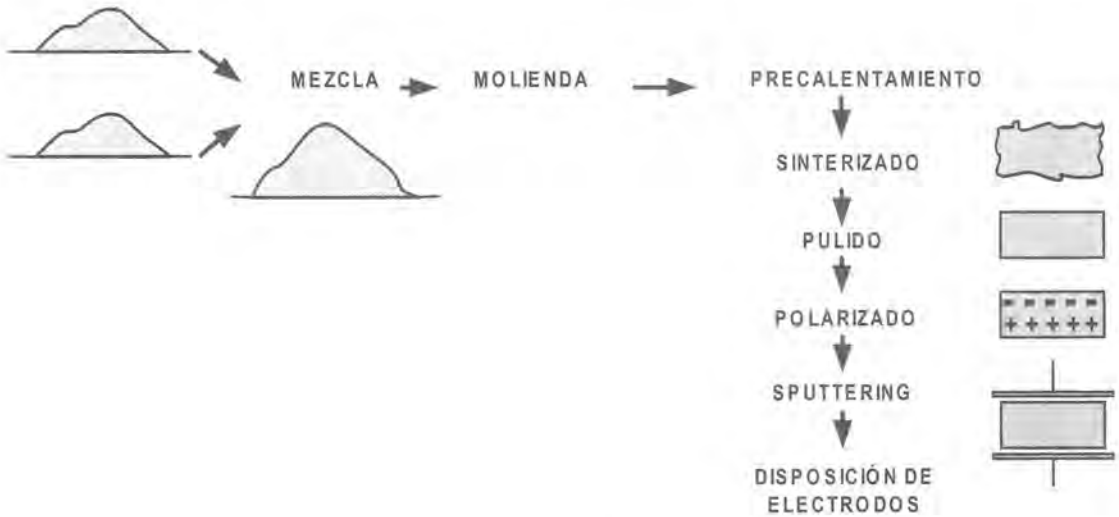


Figura 14.4. Proceso de fabricación de un sensor piezoeléctrico.

Al concluir el proceso de fabricación se dispone de un cristal de material piezoeléctrico con dos electrodos capaz de producir una carga sobre ellos cuando está sometido a un esfuerzo de deformación o de actuar en sentido contrario, es decir, producir una fuerza cuando se actúe eléctricamente sobre él, ya que el fenómeno es totalmente reversible. En la figura 14.5 se muestra un dispositivo de este tipo junto con su símbolo de circuito.

Pero el efecto piezoeléctrico no se restringe a unos pocos compuestos naturales o a algunas sustancias sintéticas, sino que es un fenómeno que aparece en más materiales. En los años sesenta se descubrió efecto piezoeléctrico en huesos y tendones de ballena lo que originó la apertura de investigaciones en el campo de los dispositivos piezoeléctricos orgánicos; los primeros frutos se obtuvieron en 1969 con la obtención de un comportamiento piezoeléctrico en polímeros como el PVDF que tenían, además, comportamiento piroeléc-

trico (véase siguiente capítulo). Este tipo de sustancias ha dado lugar a sensores piezoeléctricos peliculares y de diversas formas con buenas propiedades para el uso en sistemas de SONAR.

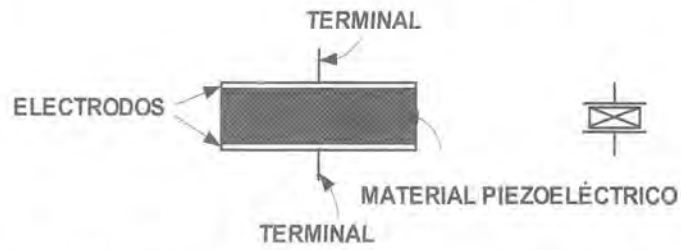


Figura 14.5. Dispositivo piezoeléctrico y símbolo de circuito.

14.2 Comportamiento de los materiales piezoeléctricos

Supongamos un material piezoeléctrico configurado como se muestra en la figura 14.5 al que se le aplica una fuerza F, provocando la consiguiente deformación (figura 14.6).

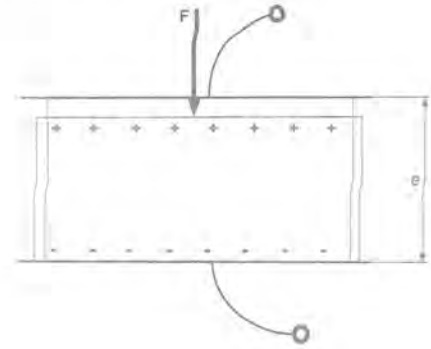


Figura 14.6. La fuerza aplicada sobre el cristal piezoeléctrico provoca la aparición de cargas.

La suma de los comportamientos inercial, elástico y viscoso del material provoca la aparición de fuerzas que contrarrestan el efecto de la fuerza F de tal manera que se puede escribir que:

$$F = m\frac{d^2x}{dt^2} + r\frac{dx}{dt} + sx \qquad (14.1)$$

siendo m la masa del cristal, r, el coeficiente de rozamiento (viscosidad), s el coeficiente elástico y x la deformación total producida. El efecto piezoeléctrico provoca la aparición de una carga directamente proporcional a la deformación producida e inversamente proporcional al espesor e del cristal:

$$q = \frac{k'}{e}x \qquad (14.2)$$

Si los terminales del dispositivo se cierran sobre un circuito, podría aparecer una corriente eléctrica según:

$$i = \frac{dq}{dt} = \frac{k'}{e}\frac{dx}{dt} = K\frac{dx}{dt} \tag{14.3}$$

lo que significa que la corriente que aparecería sería proporcional a la velocidad de variación de la deformación y que, con una fuerza estática aplicada, la deformación sería constante y la corriente tendería a anularse.

Considerando las expresiones (14.1) y (14.3), se puede escribir:

$$F = \frac{m}{K}\frac{di}{dt} + \frac{r}{K}i + \frac{s}{K}\int i dt \tag{14.4}$$

Esta expresión tiene la misma forma que la ecuación que rige la evolución de las variables eléctricas de un circuito RLC serie[3] lo que permite establecer una analogía con aquel circuito y permite una representación como la de la figura 14.7.

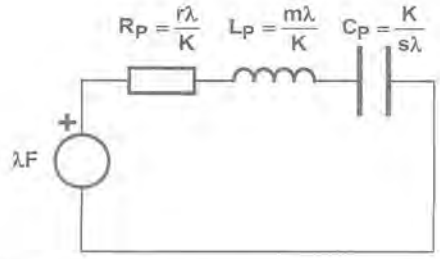


Figura 14.7. Circuito eléctrico análogo al comportamiento de un sensor piezoeléctrico.

La ecuación que representa al circuito será pues:

$$V = \lambda F = R_p i + L_p\frac{di}{dt} + \frac{1}{C_p}\int i dt \tag{14.5}$$

donde λ es un factor de conversión dimensional.

Este circuito representa perfectamente el comportamiento del dispositivo piezoeléctrico pero podemos preguntarnos: ¿dónde están los terminales de salida? Dado el aspecto del dispositivo, un par de placas-electrodos separadas por un material aislante podríamos pensar que el propio componente es el condensador del circuito de la figura 14.7 pero, en realidad no es así. En efecto, si se considera la capacidad real que representa el propio componente, asumiendo que tiene una superficie A, que los electrodos están separados una distancia e y que la permitividad dieléctrica absoluta del cuarzo es ε, se obtiene:

$$C = \varepsilon\frac{A}{e} \tag{14.6}$$

La capacidad C que corresponde al valor de la capacidad física del dispositivo, no tiene nada que ver con el efecto piezoeléctrico y es mayor que la capacidad de su circuito eléctrico equivalente. Por tanto, se considerará que la capacidad piezoeléctrica, C_P se divide en C_1 y C_2 según se indica en el circuito de la figura 14.8.

[3] Un circuito RLC serie que se excita con una fuente de tensión de valor V, está representado por una expresión como ésta:

$$V = L\frac{di}{dt} + Ri + \frac{1}{C}\int i dt$$

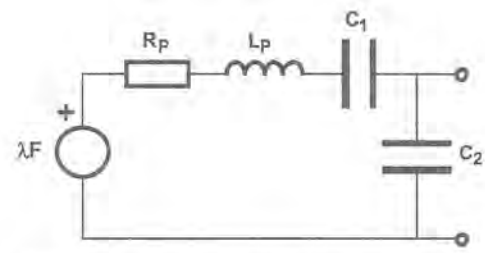


Figura 14.8. Circuito equivalente modificado del sensor piezoeléctrico.

El equivalente serie de las capacidades C_1 y C_2 forma la capacidad piezoeléctrica C_P mientras que su equivalente paralelo representa la capacidad que se ve desde el exterior del dispositivo, es decir, la capacidad construida C. Por ello, se puede escribir que:

$$\frac{1}{C_P} = \frac{1}{C_1} + \frac{1}{C_2}$$

$$C = C_1 + C_2$$

Operando, se puede obtener que:

$$C_2 = \frac{C + \sqrt{C^2 - 4\frac{k}{\lambda s}C}}{2} \qquad C_1 = \frac{C - \sqrt{C^2 - 4\frac{k}{\lambda s}C}}{2} \tag{14.7}$$

Con los valores típicos de la mayoría de los cristales se pueda considerar sin demasiado error que $C_2 >> C_1$.

Con un circuito como el que se muestra en la figura 14.8, la salida sobre la que se efectúa la medida estará cargada con una impedancia genérica Z (figura 14.9).

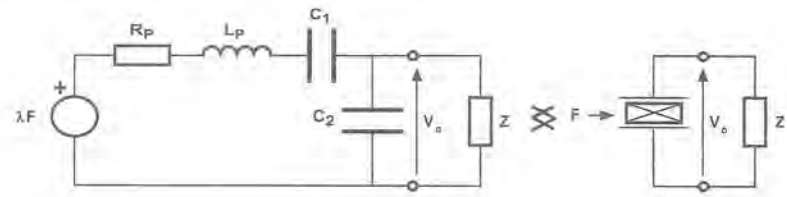


Figura 14.9. Sensor piezoeléctrico cargado con una impedancia Z.

En estas condiciones aparecen los primeros problemas:

a) Si la carga tiene una componente resistiva y la fuerza es estática, la presencia del condensador C_1 provocará que la señal de tensión en la salida vaya cayendo de valor hasta llegar a cero, lo que imposibilita al sistema para hacer medidas de esfuerzos estáticos con una salida de tensión. Visto desde otro punto de vista, si se dibuja el diagrama de Bode de la respuesta del dispositivo piezoeléctrico, aparecerá una frecuencia de corte inferior. En el fondo, lo que está ocurriendo es que se trata de un sensor generador de señal y sólo se introduce una cantidad limitada de energía por lo que no puede mantenerse indefinidamente en la salida una tensión sobre la resistencia.

b) Con cargas dinámicas —del tipo de vibraciones— se podrá desarrollar una tensión de salida de forma permanente ya que se estará aportando energía constantemente. Sin embargo, en este caso, habrá que tener en cuenta el tipo de carga ya que la presencia de componentes reactivas modificará el comportamiento del sistema (su frecuencia de resonancia y su ganancia).

Concretando, consideremos la expresión que determina la magnitud de la tensión de salida del circuito de la figura 14.8, en principio sin carga ($Z = \infty$):

$$V_0 = \lambda \; F \frac{C_P}{C_2} \frac{1}{\sqrt{\left(1-\omega^2 L_P C_P\right)^2 + \omega^2 R_P^2 C_P^2}} \tag{14.8}$$

Considerando la aproximación de que C_2 es mucho mayor que C_1 y, como quiera que el paralelo de ambos es C, se puede suponer que $C \approx C_2$. Sustituyendo las expresiones de L_P, R_P y C_P en (14.8) se obtiene la expresión:

$$V_0 = \frac{ke}{\varepsilon \; A} \frac{1}{\sqrt{\left(s-\omega^2 m\right)^2 + \omega^2 r^2}} F \tag{14.9}$$

Representado esta expresión en función de la frecuencia, se obtiene un diagrama como el de la figura 14.10, caracterizado por una ganancia en continua y una resonancia a la frecuencia f_0.

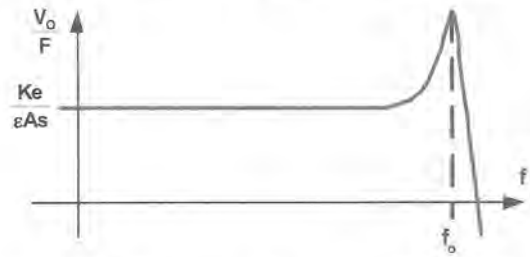


Figura 14.10. Representación de la función de transferencia de un sensor piezoeléctrico en función de la frecuencia.

La frecuencia f_0 será:

$$f_0 = \frac{1}{2\pi\sqrt{m/s}} \tag{14.10}$$

Si el circuito se considera cargado con una impedancia Z, la expresión no será la misma: por un lado, el efecto de la resistencia de carga evitará la presencia de tensión continua, mientras que la presencia de cualquier capacidad en la carga modificará la frecuencia de resonancia y la ganancia del sistema.

14.3 Los dispositivos piezoeléctricos como sensores

14.3.1 MEDIDA DE FUERZA, PRESIÓN Y ACELERACIÓN

Dada la influencia de la fuerza aplicada sobre el dispositivo piezoeléctrico en la carga generada en él, cabe la posibilidad de utilizarlo como sensor para la medida de la fuerza o de cualquier magnitud física que se pueda convertir en fuerza.

A la vista del aspecto que presenta su curva característica en función de la frecuencia (figura 14.10), siempre que se considere que está trabajando en la zona en que la respuesta es plana, se pueden despreciar los efectos inductivo y resistivo de su equivalente y, por tanto, el sistema se puede considerar reducido al circuito de la figura 4.11a en el que sólo aparecen los dos condensadores.

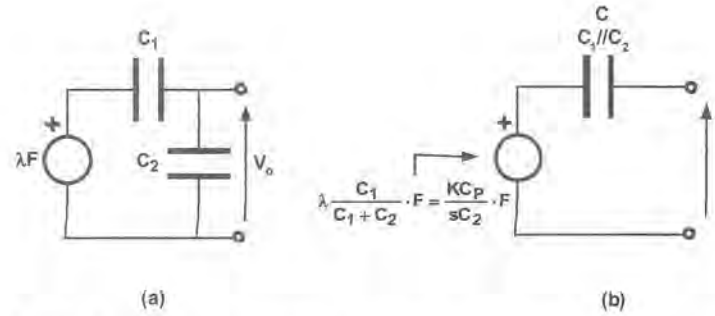


Figura 14.11. Equivalente simplificado de un sensor de fuerza piezoeléctrico.

Haciendo el equivalente de Thévenin, se obtiene el circuito de la figura 14.11b que representa el sistema con una fuente de tensión correspondiente al valor de la tensión de salida en vacío y una impedancia de salida que es el condensador que se ve entre los terminales de salida del dispositivo.

En general, los fabricantes proporcionan el valor del condensador de salida del dispositivo y la sensibilidad del sistema al parámetro que se está midiendo, suponiendo que se trabaja siempre en la zona plana de la curva de respuesta. También se suele proporcionar la frecuencia de resonancia y/o el margen de frecuencias de validez de los datos proporcionados (capacidad y sensibilidad).

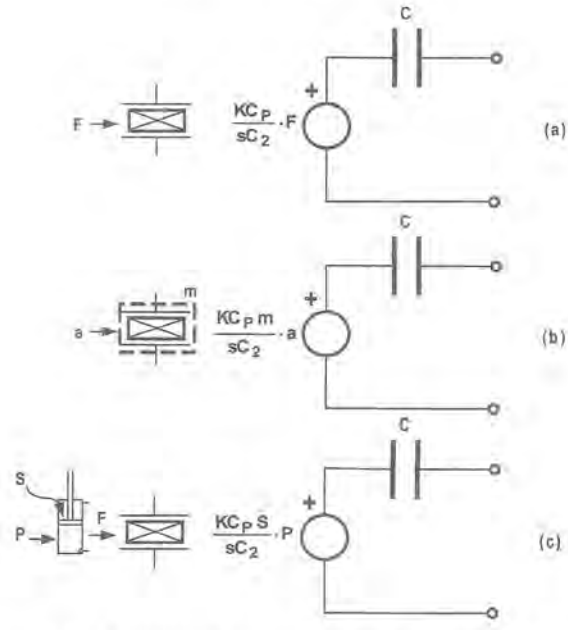


Figura 14.12. Utilización de sensores piezoeléctricos: (a) la medida de fuerza es directa; (b) la aceleración se mide con una masa m solidaria al sensor; (c) la presión se puede medir mediante un pistón de superficie S aplicado sobre el sensor que transforma la presión en fuerza.

El sensor piezoeléctrico, en sí mismo, responde a la fuerza aplicada pero, como quiera que otras magnitudes tales como la aceleración o la presión se pueden relacionar directamente con la fuerza aplicada, es posible disponer también de sensores para la medida de estas dos magnitudes; es más, los fabricantes suelen utilizar una serie de materiales piezoeléctricos que mediante la adición de masas o la determinación de una superficie de actuación pueden actuar como sensores de aceleración o como sensores de presión ya que:

$$\text{Añadiendo una masa m} \rightarrow F = ma$$

$$\text{Actuando sobre una superficie S} \rightarrow F = PS$$

En la figura 14.12 se muestran los circuitos equivalentes para los sensores piezoeléctricos de fuerza, presión y aceleración, indicando la sensibilidad del sistema a cada una de estas variables en función de los parámetros del sistema.

14.3.2 PROBLEMÁTICA DE LA UTILIZACIÓN DE SENSORES PIEZOELÉCTRICOS

La utilización de los circuitos de la figura 14.12 no es tan inmediata como pueda parecer de un circuito tan sencillo como el equivalente mostrado. En efecto, si se pudiera extraer la información del dispositivo mediante la lectura de la señal con algún sistema de impedancia infinita, es decir, sin transferencia de energía hacia el sistema de lectura, todo lo dicho hasta este momento sería totalmente cierto y la medida resultaría muy sencilla.

Desgraciadamente, cualquier sistema para la lectura del valor de salida del sensor no posee impedancia infinita por lo que cargará inevitablemente el circuito. En estas circunstancias, la carga colocada en la salida modificará el comportamiento del sistema, introduciendo un polo a baja frecuencia y estableciendo la correspondiente frecuencia de corte inferior según se muestra en la figura 14.13.

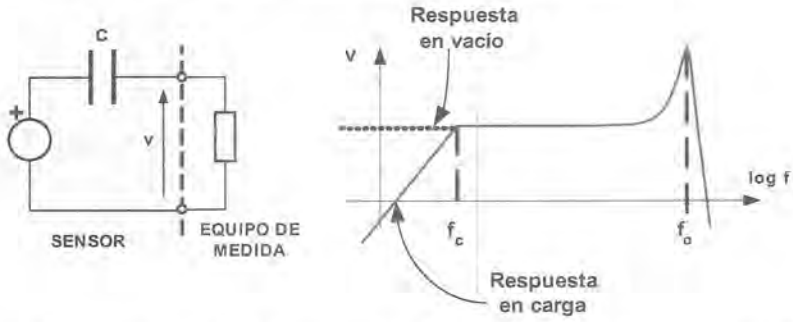


Figura 14.13. Aparición de una frecuencia de corte inferior en la respuesta del sensor piezoelécrico debido a la carga conectada.

Los efectos de la resistencia de salida sobre el comportamiento del sensor se traducen en la reducción del margen de frecuencias al que se puede usar; a pesar de que puede hacerse que la frecuencia de corte inferior sea muy baja, no puede llegar a medir en continua por lo que el ámbito de aplicación de este tipo de sensores queda limitado —en principio— a las medidas dinámicas, es decir, medidas en las que la fuerza aplicada sobre el dispositivo (la aceleración o la presión en su caso) sea variable.

Medidas dinámicas

Para los casos de las medidas dinámicas se puede hablar de una zona de trabajo comprendida entre la frecuencia de corte inferior y la frecuencia de resonancia. Sin embargo, a la frecuencia de corte inferior la respuesta del sistema es –3 dB menor que en la zona plana y, además, habrá un desfase claro entre la fuerza de

excitación y la señal entregada. En la parte superior, tampoco se deberá trabajar en las cercanías de la frecuencia de resonancia ya que se introducirían ganancias adicionales y desfases importantes por lo que se recomienda un valor mucho menor tal como $9/40f_0$ [14.6], [14.7] que introduce un error inferior al 5%. Si se pretende minimizar el error en las frecuencias extremas de la zona de trabajo, se puede trabajar una década por encima de la frecuencia de corte inferior y una década por debajo de la frecuencia de resonancia, consiguiendo que el error en la medida sea de un 1% aproximadamente (figura 14.14).

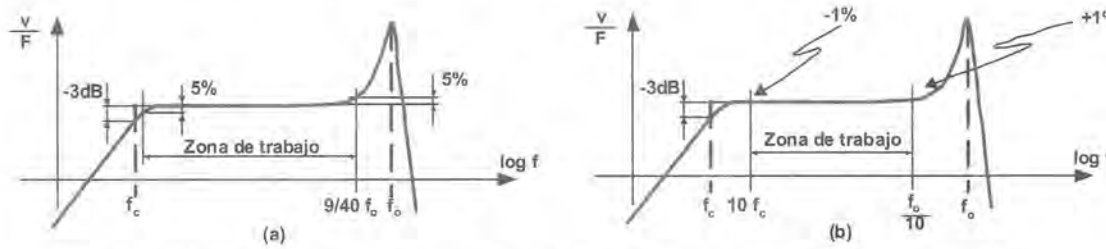


Figura 14.14. (a) Definición "tradicional" de la zona de trabajo y, (b) redefinición de la zona de trabajo para que el error sea menor o igual al 1%.

Pero aunque se tengan en cuenta las anteriores circunstancias, la medida puede traer más complicaciones ya que lo dicho hasta aquí presupone condiciones pseudo-ideales en la conexión de un equipo de medida al

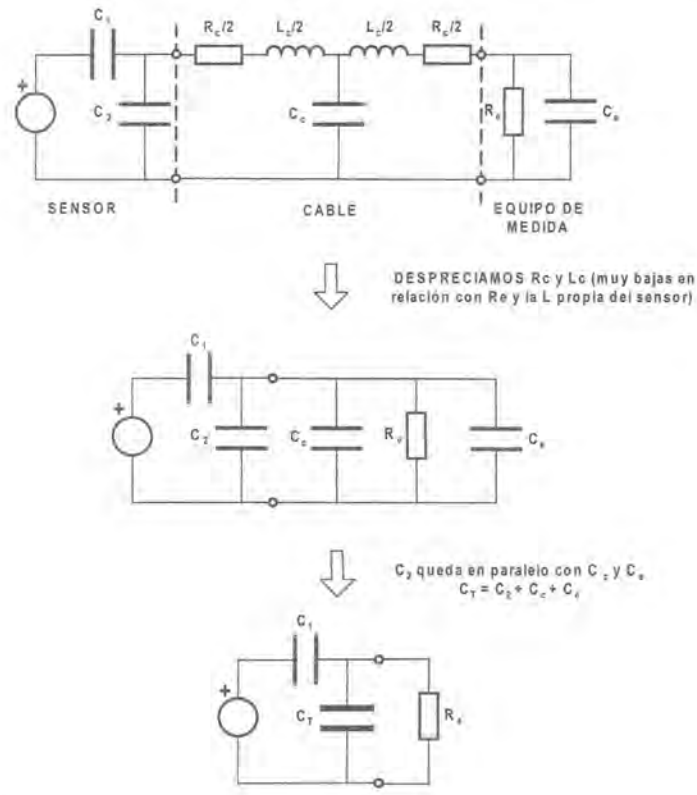


Figura 14.15. Efecto del cable sobre el circuito del sensor piezoeléctrico.

sensor (sólo se ha tenido en cuenta la resistencia de carga). Además, será necesario utilizar algún tipo de cable de conexión que introducirá un equivalente resistivo-capacitivo modificando sustancialmente la medida.

Si consideramos el equivalente de la figura 14.12a encontraremos que cualquier capacidad introducida por el cable (o la que introduzca la entrada del equipo de medida) quedará en paralelo con C_2, afectando al divisor capacitivo y, por ende, afectando a la tensión de salida según se muestra en la figura 14.15.

Como se puede observar el equivalente final tiene un aspecto igual al original excepto en el valor de la capacidad C_2 que ahora es C_T y que engloba, además, el valor de la capacidad del cable C_c y la capacidad de entrada C_e.

$$C_T = C_2 + C_C + C_e \tag{14.11}$$

Este cambio en el valor de las capacidades establece una modificación en la respuesta del sistema ya que la zona plana se desplaza hacia abajo (menor sensibilidad del sistema) y la frecuencia de resonancia baja de valor según se indica en la figura 14.16.

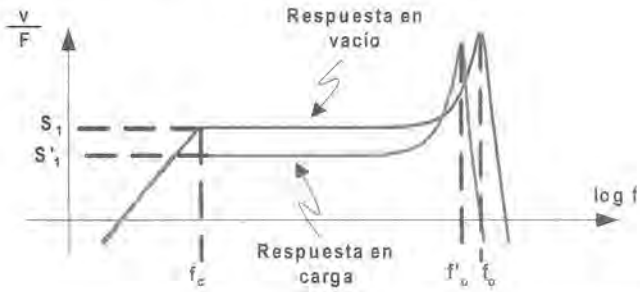


Figura 14.16. La presencia del cable de conexión modifica la ganancia y la frecuencia de resonancia del conjunto.

Pero los problemas no terminan aquí; la capacidad introducida por el cable y por el amplificador restringen las frecuencias de uso del sistema, aunque el principal problema está en que la capacidad total afecta a la ganancia en la zona de uso (la región plana). Para un sistema de medida determinado (C_e conocida) y para un cable determinado (C_C conocida) el valor de la sensibilidad S_F en la zona plana es:

$$S_F = \frac{k}{s}\frac{C_P}{C_T} \tag{14.12}$$

Por lo que cualquier cambio en el valor de C_T le afecta directamente. No es frecuente un cambio en C_e pero sí que puede ocurrir un cambio en la capacidad del cable si modificamos su longitud ya que el valor de C_C se puede obtener como la capacidad propia del cable por unidad de longitud multiplicada por la longitud empleada. Así pues, cualquier cambio en este sentido se traducirá inmediatamente en un cambio en la zona de uso y, lo que es peor, un cambio en la tensión de salida del sistema.

Finalmente, el cable introduce un factor que reduce la sensibilidad del sistema a medida que se incrementa la longitud; como quiera que los valores de sensibilidad de estos sistemas —como la de la mayoría de los sensores generadores de señal— no es muy alta, la longitud del cable debe ser minimizada para conseguir reducir C_C y, por ello, C_T.

Como precaución adicional, cuando el sensor piezoeléctrico se use para medir vibraciones, el cable debe sujetarse firmemente para evitar la introducción de ruido en la conexión (efecto triboeléctrico) lo que es especialmente importante en caminos de alta impedancia [14.12]. En la figura 14.17 se muestra la forma típica de sujeción del cable de estos sensores mediante una grapa o un adhesivo.

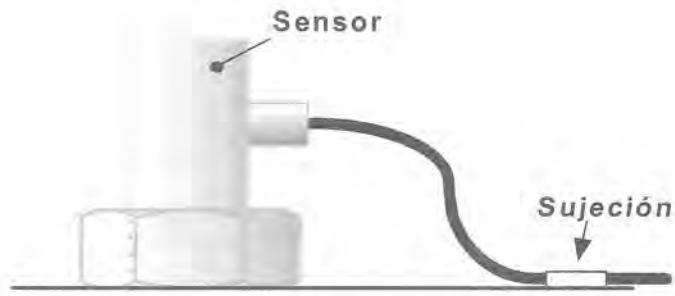


Figura 14.17. Precauciones de montaje de un sensor piezoeléctrico midiendo vibraciones.

A la hora de utilizar este tipo de sensores en medidas dinámicas ya sean de fuerza, aceleración o presión nos podemos encontrar con diversos tipos en el mercado con distintos valores de sensibilidad, alcance de la medida, banda de frecuencias de uso, etc. aunque la mayoría suelen ser de dos tipos, los sensores propiamente dichos y los que incorporan un amplificador:

- Los sensores piezoeléctricos pre-amplificados van siendo cada vez más habituales por la comodidad de uso, ya que producen un valor de tensión proporcional a la excitación aplicada en la salida del amplificador y su comportamiento resulta independiente del conexionado exterior puesto que carga y resistencia de entrada del amplificador se mantiene constantes siempre. El fabricante de estos sensores suele proporcionar como datos básicos la sensibilidad en tensión, la banda de frecuencias de uso y la impedancia de salida; para utilizarlos se puede sólo emplear su equivalente Thévenin de salida como cualquier circuito electrónico (figura 14.18). Lógicamente, este tipo de sensores precisa alimentación.

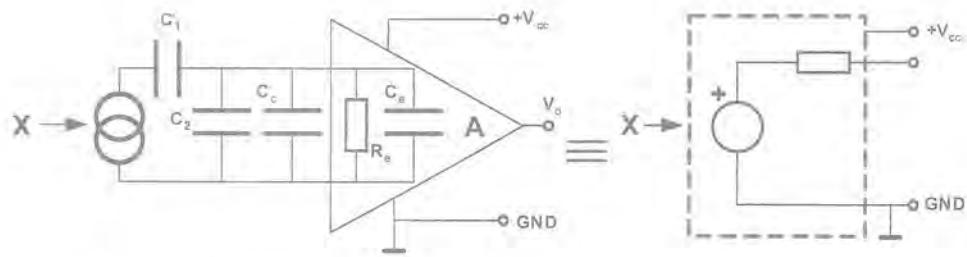


Figura 14.18. Sensor piezoeléctrico pre-amplificado y su sencillo equivalente.

- Los sensores piezoeléctricos propiamente dichos no incorporan más que el dispositivo sensor, careciendo de una salida tan cómoda como los anteriores. En este caso, todas las precauciones mencionadas antes son imprescindibles.

  Además de los datos de los márgenes de uso típicos, los fabricantes proporcionan los datos de la capacidad de salida y de la sensibilidad para la carga, es decir, la caracterización de un equivalente similar al que se mostraba en la figura 14.11b pero en el que se ha utilizado un equivalente eléctrico distinto con una fuente de carga en lugar de una fuente de tensión. En la figura 14.19 se indica este nuevo equivalente —el que más habitualmente proporcionan los fabricantes— en relación con el anterior [14.2].

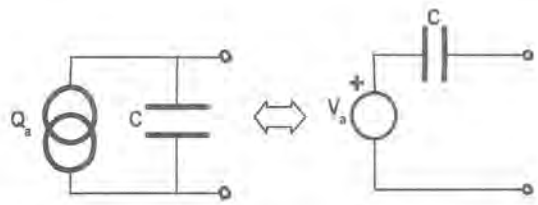


Figura 14.19. Equivalente eléctrico usual de un sensor piezoeléctrico.

Para hacer la conversión entre uno y otro equivalente basta con realizar un paso similar al que se hace para convertir un equivalente Thévenin en uno Norton o viceversa.

EJERCICIO 14.1

Se pretende hacer una medida de aceleración dinámica utilizando un sensor piezoeléctrico con una sensibilidad de 0,136 pC/ms^{-2} y una capacidad de 650 pF capaz de trabajar entre 1 y 12000 Hz con un error menor del 5%. El cable de medida tiene una capacidad de 100 pF/m y la longitud necesaria para conectarlo al amplificador es de 30 cm. Si el amplificador está constituido por una etapa con un único operacional y tiene una capacidad de entrada de 12 pF, una resistencia de entrada de 10^5 MΩ y una ganancia de 2000, determine la salida del amplificador cuando el sensor es sometido a una aceleración de 0,5 sen(1000t) m/s^2.

¿Qué limitaciones impone el amplificador operacional en el comportamiento del sistema?

SOLUCIÓN:

En la figura 14.20 se muestra el esquema eléctrico básico de esta conexión con los valores que corresponden a la aplicación específica con el valor de *100 pF/m x 0,3 m = 30 pF para la capacidad del cable.*

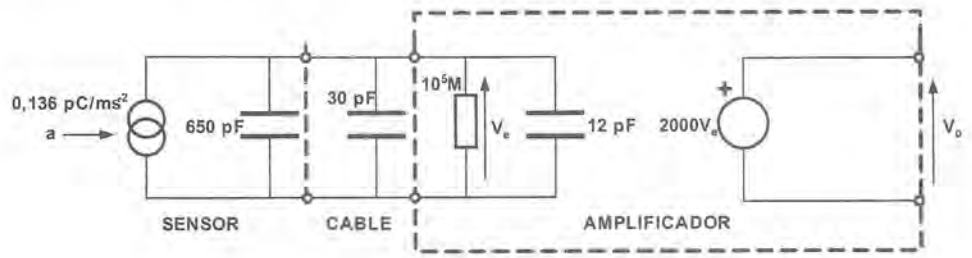


Figura 14.20

La capacidad total que aparece en la salida es, según (14.11), la suma de todas las capacidades por lo que su valor será de 692 pF. El valor de la resistencia de entrada es muy alto y puede despreciarse sin grandes errores por lo que el esquema de la figura 14.20 se puede transformar en el de la figura 14.21.

La tensión de salida del sensor que corresponde con la tensión de entrada del amplificador y que incluye el efecto de las diversas capacidades presentes será:

$$V_e = \frac{Q}{C_T} \Rightarrow V_e = \frac{0{,}136\ \text{pC/ms}^{-2}}{692\ \text{pF}} = 0{,}196\ \text{mV/ms}^{-2}$$

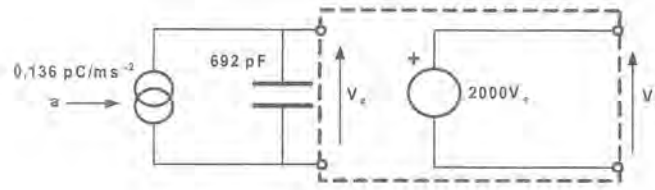


Figura 14.21

Con la aceleración presente en la entrada, la tensión de salida del sensor será $V_e = 0{,}098\text{sen}(1000t)$ mV y, tras la ganancia impuesta por el amplificador, la tensión de salida del amplificador será de $V_o = 196\text{sen}(1000t)$ mV.

Para que los valores anteriores sean ciertos, es necesario que el sensor se encuentre en su zona de respuesta plana y que el amplificador utilizado tenga un comportamiento real adecuado. En cuanto al primer aspecto, la presencia de más capacidades que la propia del sensor va a modificar su comportamiento frecuencial, desplazando la frecuencia de resonancia a menores valores. No obstante, el incremento de capacidad es muy pequeño (apenas un 6,5%) por lo que el cambio en dicha frecuencia será menor del 3,1%; como quiera que estamos trabajando en una zona suficientemente alejada de los extremos de la banda útil, no aparecerá ningún efecto adicional al del propio cambio de ganancia que ya hemos tenido en cuenta al calcular V_e.

En lo que se refiere al amplificador operacional, debe trabajar a una frecuencia de 1000 rad/s = $1000/2\pi$ Hz y con una ganancia de 2000 por lo que es preciso que su producto ganancia-ancho de banda sea de, al menos, 318 kHz lo que es muy sencillo de conseguir con cualquier operacional. El *slew-rate* no resulta demasiado crítico en esta primera etapa de amplificación habida cuenta del bajo nivel de la tensión de salida y que sólo impone un valor mínimo de 196 V/s, ridículo para cualquier operacional.

Los errores introducidos en continua pueden llegar a ser peligrosos para el correcto funcionamiento del equipo dado que, con una ganancia elevada, el nivel de salida podría llegar a saturar la salida del operacional según se vio en el capítulo 2.

Medidas estáticas

Lo explicado hasta este momento parece descartar este tipo de dispositivos cuando se pretende hacer una medida estática en la que la fuerza aplicada (o la presión o la aceleración) tiene un valor constante en el tiempo: a la vista de las figuras 14.13, 14.14 y 14.16 la sensibilidad del dispositivo piezoeléctrico a frecuencia cero es nula.

La causa está en que la generación neta de carga en el sensor es proporcional a la fuerza aplicada. Si se aplica una resistencia en la salida, la energía asociada a esta carga se disipará en la resistencia durante un corto lapso de tiempo (según las constantes del circuito) pasado el cual, no aparecerá señal. Esto tiene sentido físico ya que la fuerza aplicada ocasionará un desplazamiento determinado que se traduce en energía y, una vez disipada esta energía, como la fuerza no provoca nuevos desplazamientos, no ocasiona trabajo y no hay nuevos aportes de energía: la salida del sensor es nula y, en consecuencia, la medida de fuerzas estáticas no produce una señal continua en la salida. La representación física de este efecto se tiene mediante el condensador en serie con la fuente de tensión del equivalente eléctrico del circuito.

Sin embargo, aun en estas circunstancias, es posible hacer una medida estática bajo determinadas condiciones. Consideremos el equivalente de carga del circuito en el que la fuente de tensión se ha cambiado por

la fuente de carga equivalente (figura 14.19): como quiera que la cantidad de energía que se produce es la ocasionada durante la aplicación de la fuerza sobre el sensor y no hay nuevos aportes, si la carga producida se pudiera almacenar en algún dispositivo externo al sensor podría ser leída y, consecuentemente, informar sobre el valor de la fuerza (o la aceleración o la presión) aplicada. La realización práctica de esta idea se materializa mediante el denominado **amplificador de carga**, un circuito capaz de extraer la carga eléctrica producida en el sensor y almacenarla en un condensador externo; en la figura 14.22a se muestra el circuito básico de este sistema.

En la figura 14.22b se muestra cómo, cuando se aplica una fuerza constante F sobre el sensor y se produce la correspondiente carga eléctrica Q, esta desarrolla una tensión sobre el condensador C_A del circuito amplificador que aparece directamente en la salida del mismo con signo contrario; en esta figura se ha suprimido la capacidad del propio sensor ya que se encuentra cortocircuitada a masa virtualmente mediante la entrada del operacional. En condiciones ideales, la tensión se mantendría constante a la salida del operacional siempre que no cambiase la fuerza aplicada sobre el sensor, incluso aunque se conectase una carga ya que sería el propio operacional el que proporcionaría la potencia para mantener la tensión de salida constante.

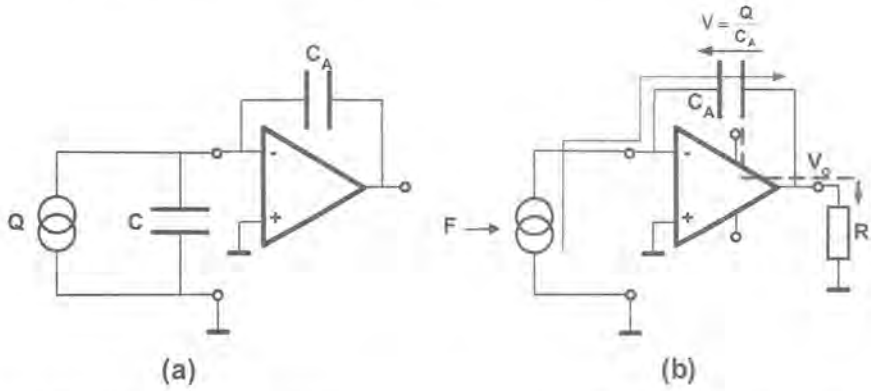


Figura 14.22. Amplificador de carga para medidas estáticas con un sensor piezoeléctrico.

En estas condiciones, la tensión de salida del operacional viene dada por:

$$V_o = -\frac{Q}{C_A} = -\frac{kF}{C_A} \tag{14.13}$$

siendo k la sensibilidad del sensor expresada en C/N. El nombre del circuito es el de amplificador de carga puesto que proporciona una tensión de salida constante proporcional a la carga de entrada fija que se produce en su entrada.

Este funcionamiento básico está limitado gravemente por las no-idealidades del operacional hasta el punto de que el circuito de la figura 14.22 no puede funcionar en condiciones reales por su topología que presenta problemas iguales a los de otros circuitos similares, como los integradores que se han tratado en el capítulo 3 y los circuitos de acondicionamiento para fotodiodos que se tratarán en el capítulo 16.

El primer problema que presenta el circuito de la figura 14.22 se debe al comportamiento en continua del operacional y que se manifiesta en las tensiones de desviación y en las corrientes de polarización que ocasionarán una rápida deriva de la tensión en el condensador por integración de los errores hasta llevar el dispositivo a saturación, eliminando la capacidad de amplificar la carga. Así, el circuito estaría saturado antes de que la fuerza actuase sobre el sensor, haciendo que la salida del amplificador no siguiese a la señal de entrada. El efecto producido es el mismo que en el caso de los circuitos integradores, pero la solución no

puede ser la misma ya que la utilización de una resistencia en paralelo con el condensador C_A causaría la disipación sobre ella de la energía de la carga eléctrica generada, causando la caída de la tensión de salida a cero en poco tiempo.

El problema puede resolverse garantizando que el condensador C_A esté descargado antes de hacer la medida y procurando que no se cargue con los errores de continua durante el tiempo que dure la medida. Para lograrlo se emplea un contacto gobernado que cortocircuita el condensador de forma permanente excepto durante el tiempo de medida y un operacional de buenas características en continua que pueda garantizar una ventana mínima de medida antes de que la integración de los errores de continua invalide esa medida. En la figura 14.23a se muestra el circuito propuesto, mientras que las formas de onda características se presentan en la figura 14.23b.

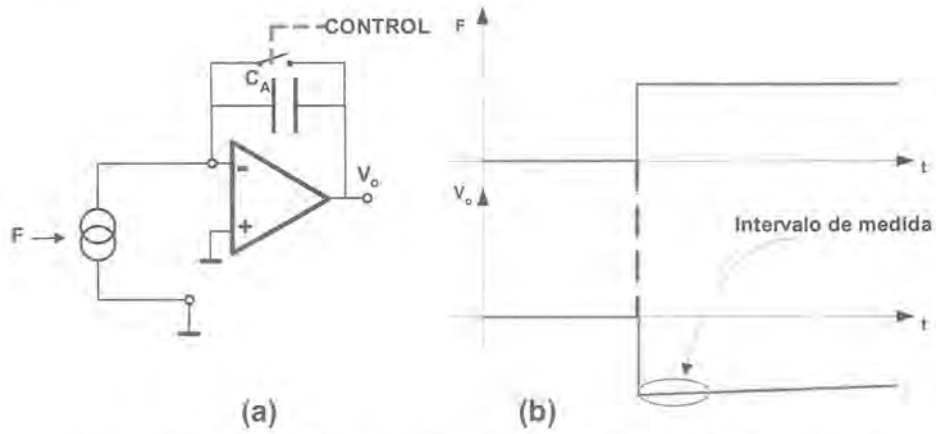


Figura 14.23. Circuito para medidas estáticas con un sensor piezoeléctrico.

El funcionamiento del circuito exige la presencia de un circuito de control que actúe sobre el interruptor y que sea capaz de leer la tensión de salida durante el tiempo que mantiene abierta la ventana de medida. Aun en este caso, la tensión contiene errores ya que el circuito integra los errores de continua del operacional durante todo el tiempo de medida. Por ello, la exactitud de la medida será mayor en la medida en que la ventana sea más pequeña y que los errores del operacional sean menores.

El segundo problema que aparece es la falta de limitación de la corriente que entrega el sensor y que cargaría en tiempo nulo el condensador; en realidad esto no es así puesto que el sensor tiene una resistencia serie según hemos visto y el propio operacional limita la corriente que puede circular por el condensador; no obstante, aun esta corriente podría ser excesiva y suele incluirse una resistencia en serie con la salida del sensor (figura 14.24) que no modifica el valor de la medida aunque introduce un retraso en la carga del condensador que debe tenerse en cuenta a la hora de efectuar la lectura de la tensión de salida del amplificador.

Finalmente, habrá que tener precauciones con el comportamiento frecuencial del amplificador operacional para garantizar que el circuito no se convierta en inestable debido a la red de realimentación utilizada.

Como única ventaja respecto de las medidas dinámicas, hay que tener en cuenta que las medidas estáticas no se ven afectadas por el cable utilizado ya que sus capacidades parásitas quedan en paralelo con C y, por tanto, cortocircuitadas virtualmente a masa por el operacional; así pues, el cambio en el cableado de conexión o en su longitud no influyen directamente en la medida y no se precisa recalibrar el conjunto. La única influencia está en la introducción de parásitos resistivos e inductivos en serie con la resistencia R (figura 14.24) que introducirían retardos adicionales en la medida.

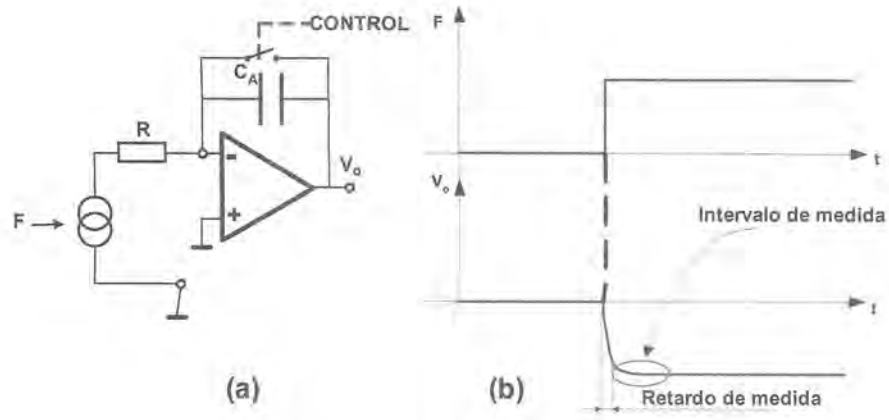


Figura 14.24. Circuito para medidas estáticas con un sensor piezoeléctrico incluyendo la limitación del pico máximo de corriente inicial.

EJERCICIO 14.2

Se dispone de un sensor de fuerza de tipo piezoeléctrico con una sensibilidad de 5 pC/N, cuya capacidad de salida es de 1200 pF y que posee un alcance en la medida de 1000 N. Se pretende realizar una medida estática con un amplificador de carga similar al propuesto en la figura 14.22 y realizado con un condensador de 1000 pF sobre un operacional que posee una corriente de polarización de 10 pA mientras que la tensión de desviación se puede considerar compensada durante el proceso de medida. ¿Cuál es la máxima tensión de salida del conjunto cuando el sensor es sometido a una fuerza constante de 500 N? ¿Cuál es el máximo tiempo que se puede abrir la ventana de medida para que el error a fondo de escala sea menor del 2%?

SOLUCIÓN:

Una vez que el sistema de control abre el interruptor, el sistema queda listo para la medida. Si se aplica una fuerza de 500 N, la tensión de salida del sistema —en su máximo en valor absoluto— será, siguiendo la expresión (14.13):

$$V_o = -\frac{kF}{C_A} \Rightarrow V_o = -\frac{5\,^{pC}/_{N} \times 500\ N}{1000\ pF} = -2{,}5\ V$$

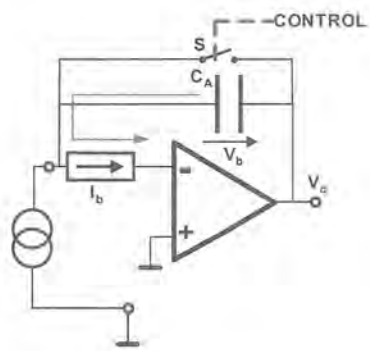


Figura 14.25

La corriente de polarización en la entrada inversora del operacional circula por el condensador según se indica en la figura 14.25, generando una tensión igual a la integral de su valor:

$$V_b = \frac{1}{C}\int i_b dt \Rightarrow V_b = \frac{1}{1000\ pF}\int_0^t 10\ pA dt = 0{,}01t$$

El fondo de escala se establece para el máximo de la fuerza de entrada, es decir, 1000 N que corresponde a un valor de salida del amplificador de 5 V; el 2% de ese valor es el máximo valor que puede tomar V_b (observe que se ha puesto un sentido arbitrario para la fuente de corriente de i_b en la figura 14.25) y que corresponde a 100 mV. Teniendo en cuenta la expresión de V_b, el máximo tiempo que se puede abrir la ventana de medida es de 10 s. En la figura 14.26 se muestran las formas de onda previstas a la salida del amplificador; si se usase un circuito como el de la figura 14.24, el aspecto sería ligeramente diferente, apareciendo al inicio un cambio gradual en la tensión de salida.

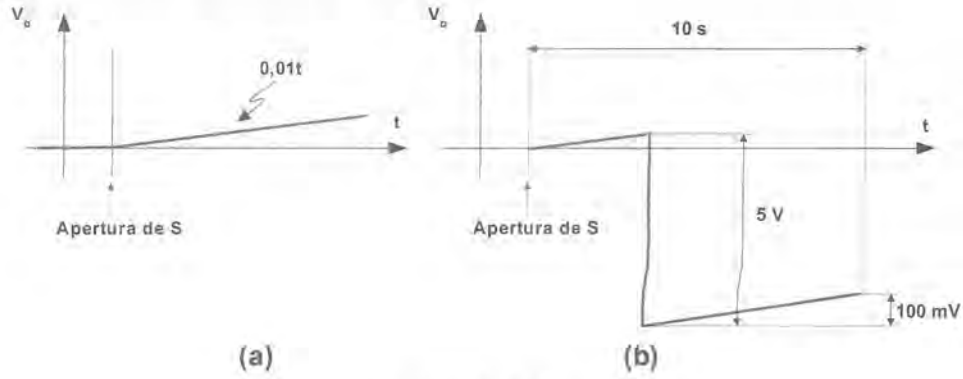


Figura 14.26

Las medidas estáticas con sensores piezoeléctricos son posibles aunque aparecen una serie de limitaciones en el uso de los circuitos de acondicionamiento que implican el uso de circuitos de control capaces de actuar sobre el sistema y de leer la señal durante un tiempo limitado que depende del error máximo que se permita; otros sistemas que permiten la medida de fuerza (peso) o presión tratados anteriormente son las galgas extensiométricas cuya principal ventaja estriba en la capacidad de medir "sin dificultad" tanto señales dinámicas como estáticas y su principal desventaja está en su baja sensibilidad.

14.4 Sensores de ultrasonidos

Algunas aplicaciones de los sensores de tipo piezoeléctrico están en el campo de los ultrasonidos en el que este tipo de dispositivos permiten captar ondas de presión en un medio elástico y generar la correspondiente señal eléctrica de salida. Dada la reversibilidad del principio de funcionamiento de los sensores piezoeléctricos, igual que ante la excitación de sonido entregan una señal eléctrica, ante una excitación eléctrica son capaces de vibrar, pudiendo provocar ondas de presión si se sitúan inmersos en un medio elástico como puede ser el aire, cualquier otro fluido o un sólido. Dada esta reversibilidad, se puede hablar de un emisor/receptor de ultrasonidos o de un transductor de ultrasonidos lo que lleva implícita la reversibilidad entre las dos variables.

La respuesta del sensor va a ser como la de cualquier sensor piezoeléctrico (figura 14.4) pudiéndose trabajar en las diversas zonas de su característica:

a) Si se pretende captar señales de ultrasonidos de cualquier origen y frecuencia, se puede realizar un funcionamiento similar al estudiado hasta ahora, aprovechando la zona plana del espectro.

b) Si se pretende utilizar la señal de ultrasonidos para medidas con la técnica impulso-eco, interesa trabajar en la zona de máxima sensibilidad para garantizar la mayor selectividad y la mejor respuesta.

Como quiera que el primer aspecto ya ha sido tratado en apartados anteriores, nos centraremos aquí en el segundo tipo de medidas que conlleva usar el sensor en el pico de resonancia, dado que en ese punto, la sensibilidad es máxima.

14.4.1 LOS ULTRASONIDOS

La definición de ultrasonidos se realiza con referencia a la captación de sonidos por el ser humano y se establece en un margen de frecuencias superiores a las frecuencias audibles. Aunque la capacidad de oír del ser humano no es una constante y, además, sufre cambios con la edad, se suele hablar de ultrasonidos cuando se trata de vibraciones en un medio elástico con frecuencias superiores a las de audición del ser humano cuyo *límite superior se fija en 20000 Hz.*

La figura 14.27 muestra el espectro de vibraciones en función de la frecuencia. La banda denominada "audible" se extiende desde 15 Hz a 20 kHz mientras que desde los 20 kHz hasta 500 MHz se encuentran los denominados "ultrasonidos".

Figura 14.27. Espectro de vibraciones acústicas.

Las aplicaciones de los ultrasonidos se extienden a más áreas que las de instrumentación y se puede hablar de su utilización en aplicaciones de soldadura (entre 20 y 25 kHz), el trabajo de metales durante los procesos de fabricación o la limpieza ultrasónica (entre 20 y 40 kHz). Para las aplicaciones de detección basadas en la técnica impulso-eco se puede trabajar en diversas bandas en función de las propiedades del medio de propagación; así, en el caso de las aplicaciones de SONAR se pueden usar frecuencias sónicas o ultrasónicas (desde 100 Hz hasta 1 MHz) y en aplicaciones de detección de faltas en materiales sólidos se suelen usar frecuencias algo más altas (desde 100 kHz hasta 10 MHz).

14.4.2 TÉCNICAS DE IMPULSO-ECO

Las técnicas de medida de tipo impulso-eco se usan en algunas aplicaciones de detección de presencia y el exponente más conocido es el SONAR (*SOund NAvigation and Ranging*) que se utiliza para la detección de objetos en aguas. Su principio de funcionamiento es muy sencillo: se emite un pulso de ultrasonido y se escuchan los ecos; el tiempo transcurrido entre el pulso emitido y el eco recibido multiplicado por la veloci-

dad de propagación en ese medio nos informa del doble de la distancia que separa el emisor del objeto que produjo el eco. Esta técnica se usó con fines militares en la Segunda Guerra Mundial para la detección de submarinos bien fuera con el propio nombre de SONAR o con el nombre de ASDIC[4] e influyó muy seriamente en la guerra en el mar entre los submarinos alemanes y los buques británicos y estadounidenses. Hoy en día, la técnica inicial está en desuso ya que la recepción del pulso alerta al hipotético objetivo de la presencia del emisor antes de que el primero escuche el eco, por lo que quedaría también al descubierto y su aplicación militar se restringe al uso de sonoboyas lanzadas desde helicópteros o desde aviones antisubmarinos.

Sin embargo, aunque el origen sea militar y hoy en día se prefieran otras opciones para su función inicial, en algunos campos de la instrumentación sigue usándose con un notable éxito, como en la sismografía, en el estudio de los fondos marinos, en la detección de fallos en materiales, o en medicina. En la figura 14.28 se muestra un diagrama básico de un sistema de detección de objetos.

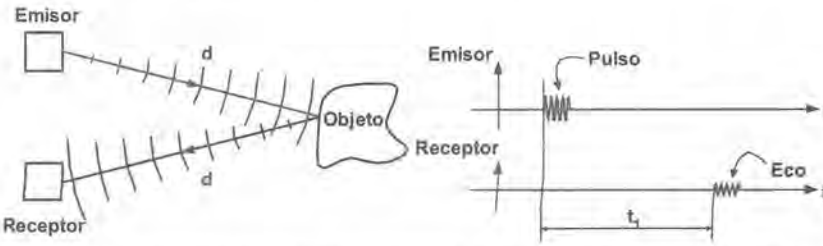


Figura 14.28. Funcionamiento básico de la técnica de medida de distancias por impulso-eco.

Si el tiempo transcurrido entre el pulso y el eco recibido es t_1 y la velocidad del sonido en el medio de propagación es c, la distancia del conjunto emisor-receptor y el objeto será:

$$d = \frac{ct_1}{2} \tag{14.14}$$

Las condiciones del medio de propagación pueden modificar su comportamiento en relación con la velocidad de propagación de las ondas de presión que constituyen el sonido por lo que el valor de c no es una constante del medio, sino que sufre variaciones según la temperatura, la presión, la densidad o la presencia de sustancias disueltas. De esta forma, se recurre a técnicas de corrección que efectúan una medida de este parámetro para después calcular la distancia al objeto. En la figura 14.29 se muestra una de esas técnicas consistente en colocar un objeto a distancia conocida del emisor por lo que se recibirán dos ecos, utilizando el primero para el cálculo de c.

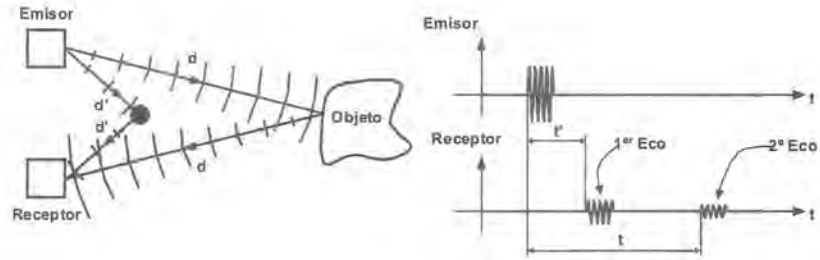


Figura 14.29. Técnica impulso-eco modificada para eliminar la incertidumbre de la velocidad de propagación.

[4] ASDIC son las siglas de *Anti-Submarine Detection Investigation Committee*.

Como quiera que d' es conocida tenemos que:

$$d' = ct' \Rightarrow c = \frac{2d'}{t'}$$

$$2d = ct \Rightarrow d = d'\frac{t}{t'}$$

Esto implica la presencia de algún dispositivo de cálculo para realizar el cociente de los dos valores de tiempo medidos además de disponer de una suficiente resolución temporal como para poder hacer la medida de t' con suficiente precisión ya que los errores introducidos en ella afectan a la medida final y podríamos llegar a tener una medida menos precisa que considerando c como una constante.

El diseño del sistema para la excitación y la lectura de la señal de eco recibida consiste en el empleo de circuitos similares a los de la figura 14.30.

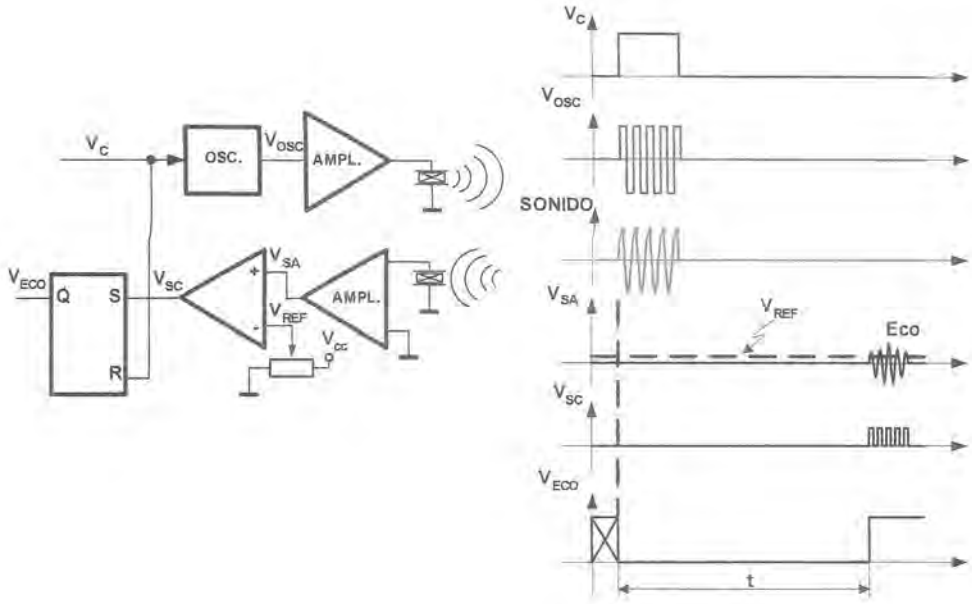


Figura 14.30. Circuito para medida de distancias por ultrasonidos mediante la técnica impulso-eco.

Cuando se activa la señal de control V_C, el oscilador produce una señal cuadrada de frecuencia fundamental la de resonancia del cristal (punto de máxima eficiencia). Esta señal es amplificada y utilizada para excitar el emisor; la onda de sonido emitida por el cristal será bastante senoidal gracias a lo restrictivo de su comportamiento en frecuencia en la zona de resonancia que eliminará con facilidad los armónicos superiores (figura 14.14). Cuando llegue el eco de la misma frecuencia de emisión[5], el dispositivo receptor entregará una señal que será amplificada y comparada con un nivel de referencia V_{REF} que establece el mínimo valor que se considerará como eco válido[6]. Superado este mínimo, la señal V_{SC} producida por el comparador activará el biestable que produce la salida de "contacto válido" denominada V_{ECO}.

[5] Sólo si se considera el objeto a detectar o el conjunto emisor/receptor en movimiento se producirían cambios en la frecuencia de la señal recibida respecto de la emitida por el efecto Doppler.

[6] El receptor podrá captar señales procedentes de ecos de objetos muy pequeños (señales muy débiles) o de otras vibraciones en el medio. La presencia de un umbral de detección permite eliminar estas señales denominadas en general como "falsos contactos".

Habida cuenta de la reversibilidad de los dispositivos piezoeléctricos cabría pensar en la posibilidad de utilizar un mismo dispositivo para las tareas de emisión y recepción, ahorrando espacio y dinero; en la figura 14.31 se ofrece una solución de este tipo.

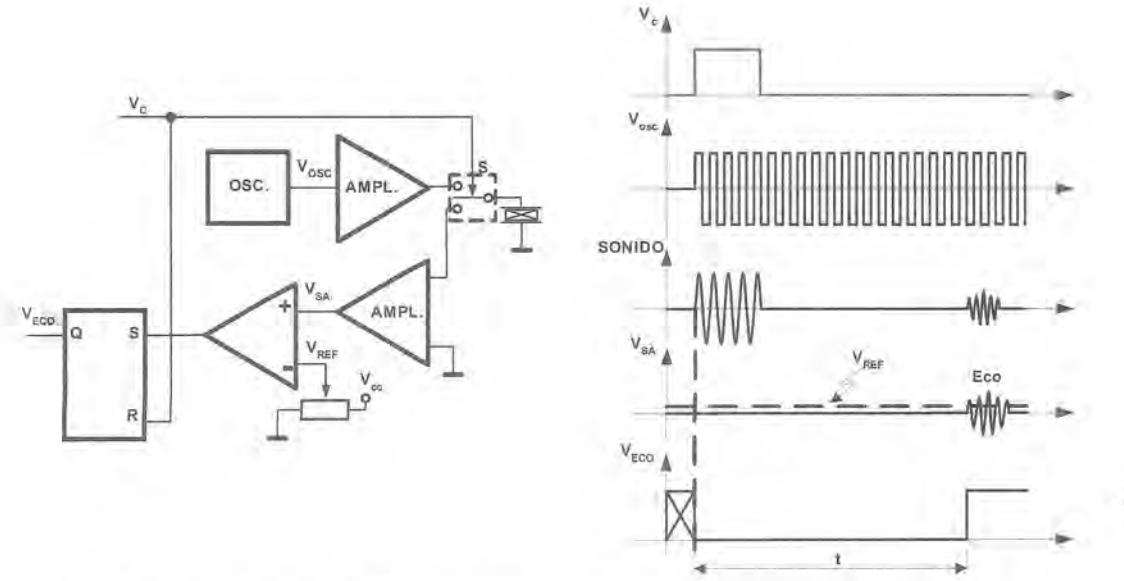


Figura 14.31. Medida de distancias empleando un único dispositivo piezoeléctrico.

Los cambios respecto del caso anterior son la inclusión de un conmutador electrónico S que conecta el cristal al dispositivo de excitación o al amplificador de recepción y que el oscilador funcionará de forma permanente aplicando su señal al cristal sólo cuando sea preciso.

Aunque el concepto de la medida no es demasiado complejo y no es necesario hacer grandes consideraciones, como las de los dispositivos piezoeléctricos en su uso como sensores, sí que hay que tener unas mínimas precauciones para garantizar un funcionamiento adecuado del sistema:

a) La frecuencia de oscilación a utilizar debe ser adecuada al cristal que se esté empleando, procurando que se encuentre situada en el punto de máxima sensibilidad definido por la frecuencia de resonancia del cristal. Separarse de ese punto se traducirá a la vez en la pérdida de potencia de emisión y en la pérdida de sensibilidad en la recepción lo que implica un sistema de menos alcance o incapaz de distinguir objetos si no tienen un tamaño suficiente.

b) Hay que diseñar con cuidado el sistema para evitar el acoplamiento del circuito de emisión con el de recepción a través de cualquier medio (interferencias capacitivas, acoplamiento a través de masas o alimentaciones, etc.) y evitar así que se dispare el receptor sin la presencia de ecos. En esta misma línea es conveniente realizar un diseño mecánico adecuado, evitando vibraciones espurias que pudiesen dar lugar a falsos contactos.

c) La actuación sobre la tensión de referencia de comparación (V_{REF}) debe hacerse con sumo cuidado ya que un valor muy alto significaría que sólo ecos de cierta potencia serían detectados, lo que se traduciría en que sólo se podrían detectar objetos muy grandes y cercanos. Por otro lado, valores muy bajos darían lugar a un gran número de falsos contactos y el sistema sería inutilizable. Para conseguir una aceptable optimización debería trabajarse con señales potentes en la emisión y receptores sensibles que proporcionen una buena posibilidad de discriminación entre lo que es y lo que no.

d) La resolución de distancias depende de la resolución temporal del equipo de medida y de la velocidad de propagación del medio: así, a mayor resolución temporal y menor velocidad de propagación, mejor será el valor de resolución de distancias.

EJERCICIO 14.3

Un sistema de medida de distancias mediante ultrasonidos en el aire está formado por un equipo de medida con una resolución temporal de 50 μs y está formado por un emisor/receptor con una frecuencia de resonancia de 35 kHz que es excitado con un tren de 8 pulsos cada vez que desea hacer una medida. Determine la resolución en distancia y obtenga la mínima distancia que es capaz de discriminar admitiendo que el objeto a detectar tiene tamaño suficiente como para generar un eco válido y que la velocidad del sonido en el aire es de 344 m/s.

SOLUCIÓN:

La resolución en distancia depende de la resolución temporal que es 50 μs; este será el tiempo mínimo que podrá detectar por lo que esta fijará la resolución en distancia que posee el sistema. A 344 m/s, la distancia que recorre el sonido en 50 μs sería de:

$$d = 344 \text{ m/s} \times 50 \times 10^{-6} \text{s} = 17{,}2 \text{ mm}$$

Como quiera que el camino se recorre de ida y vuelta, el valor de la resolución de distancia sería la mitad de ese valor, es decir, 8,6 mm.

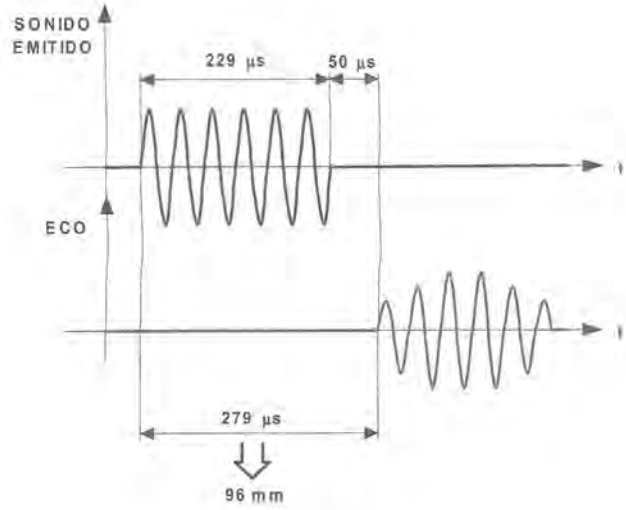


Figura 14.32

Para obtener el mínimo valor de la distancia a medir, debemos tener en cuenta el tiempo que se tarda en la emisión de la onda ya que sólo a partir de que se concluya el proceso, se abriría la ventana de escucha según se muestra en la figura 14.30. Para este caso, se envían 8 pulsos de sonido a una frecuencia de 35 kHz lo que supone un tiempo de 8/35000 = 229 μs por lo que antes de ese tiempo no se admitiría ningún eco, luego el

objeto más próximo debe estar apartado una distancia tal que el primer eco llegue como poco 229 + 50 μs después (figura 14.32), lo que significa que el límite de detección teórico del sistema será la mitad de:

$$d = 344\ \text{m/s} \times 279 \times 10^{-6}\ \text{s} = 96{,}0\ \text{mm}$$

es decir, 48,0 mm.

Hablar de la distancia máxima que un sistema de este tipo es capaz de detectar no es sencillo puesto que depende de las condiciones del medio y de la morfología y composición del objeto a detectar por lo que no resulta fácil predecir ese máximo en una configuración dada.

Otro factor a considerar en un sistema de detección de objetos por ultrasonidos es la dirección en la que se detectan los ecos que, en principio son todas. Para reducir la zona de detección lo que resulta interesante en muchas aplicaciones, se puede recurrir a la utilización de placas o conos deflectores que dirigen las ondas por un determinado lugar (figura 14.33) o a la generación de ondas planas que se propagan en una dirección determinada.

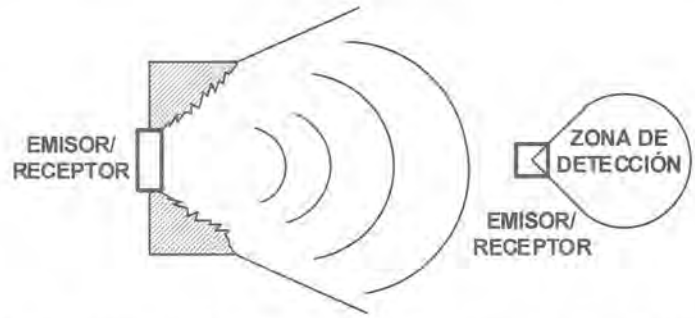


Figura 14.33. Reducción de la zona de detección mediante conos deflectores.

14.4.3 APLICACIONES DE LAS TÉCNICAS DE IMPULSO-ECO

Son muchas las aplicaciones de estas técnicas en diversos campos y engloban un gran número de posibilidades que van más allá de la detección de objetos y de la medida de distancias. A la hora de citar las aplicaciones comenzaremos por las más sencillas e iremos evolucionando hacia otras más complejas.

Detección de presencia

Es la más simple de todas y consiste en determinar si hay o no objetos presentes en las proximidades de dispositivo. En realidad, podría calcularse también la distancia pero no se considera necesario ya que la utilidad de estos sistemas está en la prevención de choques entre sistemas en movimiento o aplicaciones similares a las de otros sensores de proximidad como los fotoeléctricos. El alcance de estos sistemas es muy variable aunque sus valores típicos de uso están en distancias menores de 10 m.

Detección de fallos en materiales sólidos

Es una aplicación similar a la anterior aunque por las particularidades de la propagación del sonido en los sólidos aparecen mayores dificultades. La idea básica consiste en enviar una onda de sonido de alta frecuencia a través de un material; si aparecen grietas internas, en cada cambio de fase se producirá un fenómeno de refracción/reflexión, causando un eco que puede detectarse según se muestra en la figura 14.34.

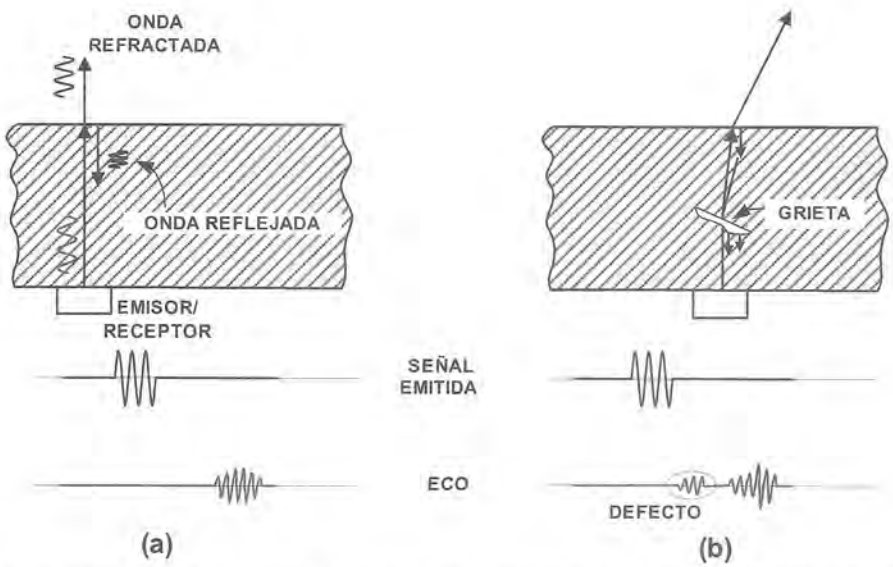


Figura 14.34. Método para la detección de fallos en materiales sólidos. (a) Material sin defectos; (b) material con defectos (grieta).

El principal problema de este sistema está en la velocidad de propagación del sonido por los materiales sólidos que, al ser muy elevada, precisa una gran resolución temporal del equipo de medida y una mayor frecuencia de resonancia en los dispositivos piezoeléctricos. Este tipo de sistemas se emplea para detectar problemas en construcciones pétreas tales como antiguos monumentos.

Otra aplicación de este tipo de sistemas está en los estudios sismográficos en donde permite detectar capas de sedimentos y, teniendo en cuenta las formas de los ecos e informaciones adicionales, permite precisar composiciones de materiales y demás características de la corteza terrestre. Este tipo de sistemas son mucho más complejos y suelen implicar un estudio adicional de datos no procedentes de los ultrasonidos.

Medida de distancias, superficies y volúmenes

La medida de distancia es la clásica aplicación de los sensores de ultrasonidos y el funcionamiento básico es el ya detallado en apartados anteriores por lo que no merece más comentarios. Sin embargo, la medida de

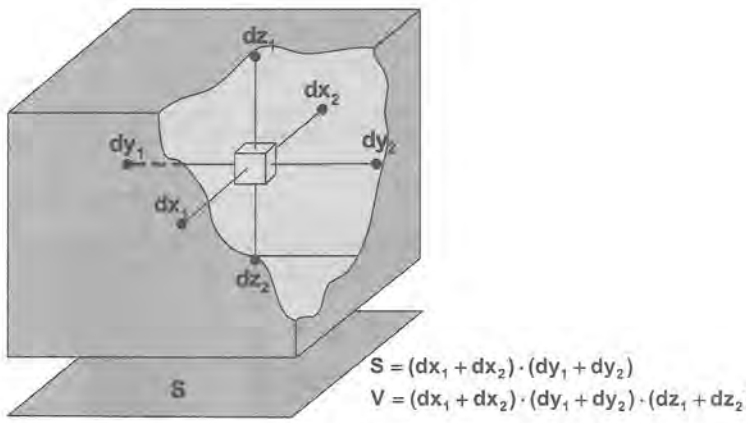


Figura 14.35. Medida de superficie y volumen interno por medida de distancias.

distancia abre otras posibilidades tales como la medida de superficies encerradas entre paredes o de volúmenes de cuerpos geométricos huecos y conocidos: suponiendo una forma de paralepípedo, el emisor/receptor se coloca en el interior y se miden las distancias en dos o tres ejes, pudiendo calcularse, en consecuencia la superficie o el volumen (figura 14.35).

SONAR

Como se dijo al principio de este capítulo, el SONAR fue el sistema que más impulsó el trabajo en sensores piezoeléctricos. Los sistemas de SONAR se pueden clasificar en dos grandes grupos:

a) Los sistemas activos: basados en la técnica impulso-eco en los que el objeto es detectado por el eco que produce ante una excitación de señal. La utilidad de este tipo de sistemas se centra en los estudios de fondos marinos, en los barcos de pesca para la detección de bancos de peces y en actividades de este tipo.

b) Los sistemas pasivos: en los que la detección de los objetos se produce por el ruido que generan. El nombre de SONAR pasivo se debe a que su aplicación es la misma que la de los primeros equipos activos basados en la técnica de impulso-eco, es decir, la detección de sumergibles por el sonido que producen.

La aplicación de los sistemas activos tiene sus peculiaridades en relación con otras aplicaciones de los sensores de ultrasonidos derivadas del medio por el que el sonido se propaga a más velocidad que en el aire y con bastante poca atenuación por lo que es factible efectuar detecciones a grandes distancias; los aspectos a tener en cuenta dependerán de la aplicación particular pero se pueden citar algunos tales como la potencia de la señal emitida, la sensibilidad del receptor, el ajuste del nivel mínimo de recepción para obtener un compromiso entre el límite de detección y los falsos contactos o la frecuencia de trabajo que puede ser determinante en algunas aplicaciones de pesca.

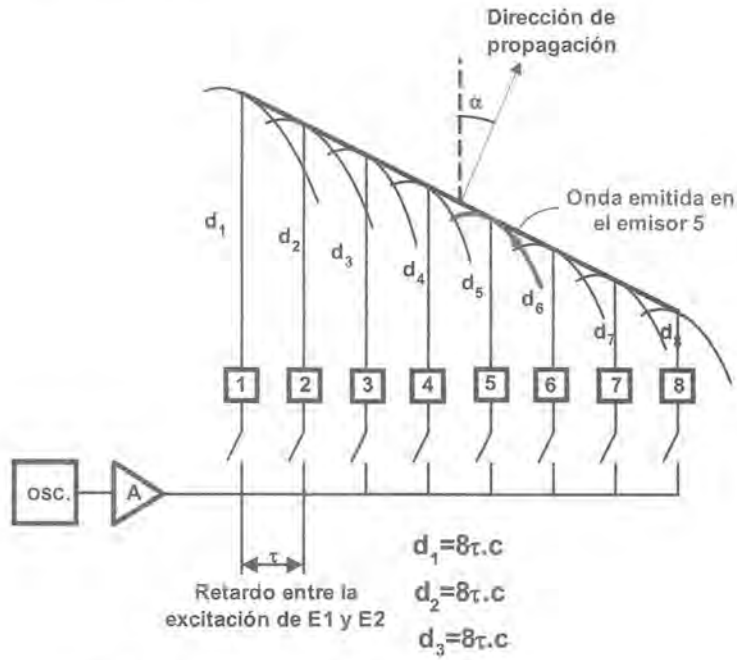


Figura 14.36. Generación de una onda de sonido plana.

Cuando se utiliza alguno de estos sistemas, el tiempo entre el pulso y el eco determina la distancia en función de la velocidad de propagación del sonido en el agua (depende de las condiciones batitermográficas[7]); pero no se tiene información acerca de la dirección del objeto que ocasionó el eco. Para conocer la dirección del eco se puede recurrir a triangulación entre dos equipos lo que suele requerir la presencia de dos barcos para tener una buena resolución de la telemetría del objeto. Otra posibilidad es generar ondas planas de sonido (mediante interferencia entre varios focos emisores) que se propaguen en una determinada dirección con lo que los ecos de intensidad vendrán de esa dirección; la obtención de ondas planas precisa un *array* de emisores, cada uno de los cuales recibe la señal de excitación con un retraso fijo respecto del anterior según se muestra en la figura 14.36.

En esta figura, el ángulo α es mayor en la medida que el retardo τ entre el momento de excitación de un emisor y del siguiente es mayor: si todos los emisores se excitan a la vez, $\alpha = 0$ y el frente de onda se desplazaría paralelo al *array* de emisores.

Sistemas médicos ecográficos

La aplicación de los ultrasonidos en medicina ha permitido realizar análisis similares a los efectuados en otros tiempos por rayos X con la ventaja de su inocuidad por lo que es aplicable a casos muy críticos como en el estudio de los procesos de gestación en los que el embrión se podría ver perjudicado por la acción de la radiación.

Los sistemas de ecografía consisten en un *array* de emisores/receptores trabajando a frecuencias altas (de 3 a 10 MHz) que reciben múltiples ecos en función de las distintas interfases (cambios de tejidos) que se vayan encontrando los pulsos de sonido en su camino. Todos estos ecos son recogidos y procesados en complejos sistemas informáticos que, finalmente proporcionan una imagen del interior del cuerpo humano.

Los últimos avances del procesamiento de información permiten la obtención de imágenes tridimensionales de gran nitidez o la medida de movimientos mediante técnicas de efecto Doppler como en los casos de análisis cardiológico o de funcionamiento del sistema cardiovascular.

Resumen

Los sensores piezoeléctricos se construyen con determinados materiales que tienen la propiedad de producir cargas cuando son sometidos a deformación. Su equivalente eléctrico resulta similar al de un circuito RLC aunque, en una primera aproximación pueden despreciarse los comportamientos resistivos e inductivos quedando sólo el comportamiento capacitivo.

La respuesta de los sensores piezoeléctricos tiene dependencia de la frecuencia de trabajo presentando un pico de resonancia, una zona plana en la que se puede trabajar y una frecuencia de corte inferior causada por la carga que se coloque en el circuito.

La utilidad de los sensores piezoeléctricos trasciende la mera medida de la fuerza ya que cualquier variable física que se pueda transformar en una fuerza puede ser medida con ellos; casos muy habituales son los de la medida de aceleraciones y vibraciones y los de medida de presión.

[7] Las condiciones batitermográficas son las condiciones de presión y temperatura sobre las aguas. Afectan a la velocidad de propagación del sonido en el agua aunque la salinidad y las corrientes también la afectan (estas últimas también tienen una buena parte de su origen en las condiciones batitermográficas).

En cualquiera de esos casos, la medida es relativamente sencilla cuando la variable a medir es una señal de alterna (medidas dinámicas), mientras que el problema se complica con señales constantes en el tiempo tales como las medidas de peso (medidas estáticas).

Cuando se pretende realizar medidas dinámicas se emplea un amplificador de alterna aunque hay que tener precauciones con los diversos parásitos de cables y entrada del amplificador, sobre todo los capacitivos ya que perturban la ganancia del conjunto y su banda de frecuencia de trabajo.

Las medidas estáticas también se pueden llevar a cabo bajo ciertas condiciones: empleando un amplificador de carga y haciendo la medida en un pequeño intervalo de tiempo aunque otras posibilidades de sensores podrían ser más convenientes para estos casos.

Un caso muy interesante de sensores piezoeléctricos es el de los sensores de ultrasonidos, dispositivos reversibles que se usan en diversas medidas basadas en la técnica de impulso-eco tales como sonar, medida de distancia, ecografías médicas, aplicaciones sismológicas, etc.

Los sensores de ultrasonidos trabajan casi siempre en el pico de resonancia para incrementar la sensibilidad y sus circuitos de medida son bastante simples de concepción: emiten un pulso y miden el tiempo que pasa hasta recoger el eco.

Problemas propuestos

14.1 Determinar la respuesta frecuencial de un sensor piezoeléctrico cuyo equivalente eléctrico incluye una resistencia de 100 Ω, una inductancia de 100 mH y dos capacidades de 0,16 y 40 pF. Dibuje también su equivalente y calcule la frecuencia de resonancia.

14.2 ¿Qué ocurre en la respuesta frecuencial si se carga la salida del sensor con una resistencia de 100 MΩ? ¿Y en la frecuencia de resonancia?

14.3 Repetir el problema 14.2 si se considera que la carga está formada por la misma resistencia y por una capacidad de 12 pF en paralelo con ella.

14.4 Un sensor de fuerza de tipo piezoeléctrico tiene una sensibilidad de 3,0 pC/N y una capacidad de 500 pF, siendo capaz de trabajar entre 10 y 10000 Hz. ¿Qué salida proporciona el sensor ante la aplicación de una fuerza constante de 20 kg?

14.5 Repita el problema 14.4 si el sensor se considera cargado con una impedancia de 1000 MΩ, dibujando la evolución temporal de la salida.

14.6 El sensor del problema 14.4 es sometido a una fuerza vibratoria de 100 Hz de frecuencia y 100 N de valor máximo. ¿Qué salida produce el sensor en vacío? ¿Y cuando está cargado con 1000 MΩ?

14.7 Extraiga conclusiones de funcionamiento del sensor según los resultados de los problemas 14.4, 14.5 y 14.6.

14.8 En el caso del sensor del problema 14.4, ¿cuál es la salida producida si se emplea un amplificador de ganancia 1000 cuyo equivalente de entrada es una resistencia de 1000 MΩ y se conecta mediante un cable de 12 pF?

14.9 Analice el cambio de ganancia en el sensor introducido por el amplificador del problema 14.8.

14.10 Un sensor piezoeléctrico de sensibilidad 0,50 pC/ms^{-2} y una capacidad de salida de 1000 pF está sometido sólo a la gravedad. ¿Qué tensión desarrolla sobre una carga de 1000 MΩ?

14.11 Si se extrae la señal mediante un amplificador de carga con un condensador de 1,0 nF, ¿cuál es la tensión de salida si el operacional es ideal? ¿Y si tiene una corriente de polarización de 1 pA?

14.12 Evalúe la forma de onda de la tensión de salida en el caso del ejercicio anterior cuando se emplea una resistencia de 5K para limitar corriente según el montaje de la figura 14.24. Determine el intervalo de medida si el error máximo permitido es del 3%.

14.13 Se desea construir un sensor de distancia por ultrasonidos para medir la distancia en el aire a objetos situados entre 1 y 20 m. Dibuje el diagrama de bloques tan detallado como sea posible siguiendo las directrices de la figura 14.30.

14.14 ¿Cómo especificaría los sensores necesarios para esta aplicación?

14.15 ¿Cómo se introduciría la corrección de la velocidad del sonido en el aire en el esquema del ejercicio 14.13?

14.16 Diseñe —a nivel de bloques— el sistema de control de un generador de ondas planas de ultrasonido compuesto por 10 emisores de ultrasonido.

14.17 Pensando en la reversibilidad de los emisores de ultrasonido del problema anterior, ¿cómo se podría hacer un sistema capaz de detectar la dirección del sonido empleando un *array* lineal de sensores de ultrasonido?

14.18 Si se pretendiese detectar la dirección del sonido procedente de una hipotética fuente de 18 Hz en agua de mar con una precisión mejor de 5º, ¿cuántos sensores (micrófonos) deberían emplearse y cuánto se deberían separar entre sí?

14.19 Si se considera la atenuación del sonido directamente proporcional al cuadrado de la distancia, esboce un sistema que permita conseguir fijar unívocamente la posición de la fuente de sonido.

14.20 Repita el problema anterior considerando la utilización de dos *arrays* de sensores separados una distancia determinada. ¿Cómo influye la distancia de separación de los *arrays* de sensores en la precisión de la determinación de la distancia?

THOMSON

2ª PARTE

INSTRUMENTACIÓN ELECTRÓNICA

INCLUYE CD

Miguel A. Pérez García
Juan C. Álvarez Antón
Juan C. Campo Rodríguez
Fco. Javier Ferrero Martín
Gustavo J. Grillo Ortega

15

Sensores piroeléctricos

15.1 Introducción
15.2 Principio de funcionamiento
15.3 Circuito equivalente de un sensor piroeléctrico
15.4 Acondicionamiento de señal
15.5 Aplicaciones

15.1 Introducción

Cualquier cuerpo a una temperatura superior a 0 kelvin (0 absoluto) emite radiación electromagnética debido a la vibración que experimentan las partículas, átomos y moléculas del cuerpo (véase apéndice 3 en el CD adjunto). Esta radiación se denomina radiación térmica. En principio existen dos tipos de sensores conocidos por su capacidad para responder a la radiación térmica: los **sensores cuánticos** y los **sensores térmicos**. Los sensores cuánticos serán tratados en el siguiente capítulo y están basados en el efecto fotoeléctrico o efecto que provoca la interacción de fotones en la red cristalina de un material semiconductor. Los sensores térmicos están basados en el incremento de temperatura que se produce en ciertos materiales al absorber parte de la radiación incidente; el incremento de la temperatura puede ser una medida de la radiación generada por la fuente. En este capítulo analizaremos el principio de funcionamiento del **sensor piroeléctrico** que es un tipo de sensor térmico basado en el denominado efecto piroeléctrico.

15.2 Principio de funcionamiento

Los sensores piroeléctricos son detectores de radiación térmica y se usan fundamentalmente como sensores de radiación en el infrarrojo (IR). Presentan una respuesta espectral bastante uniforme en un margen amplio de longitudes de onda (desde el ultravioleta (UV) hasta microondas). El efecto piroeléctrico está relacionado con el efecto piezoeléctrico estudiado en el capítulo anterior; la diferencia estriba en que ahora las cargas son inducidas fundamentalmente por efecto térmico y no mecánico. Un sensor piroeléctrico está formado por un pequeño condensador, constituido por una pastilla muy fina de material cerámico ferroeléctrico con dos electrodos en su superficie sobre los cuales se induce una carga eléctrica por efecto térmico (figura 15.1).

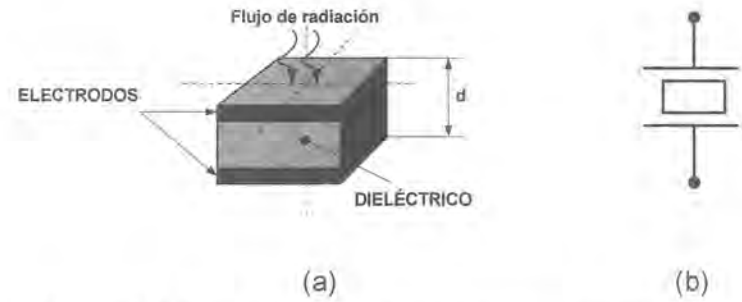


Figura 15.1. Estructura básica de un detector piroeléctrico.

El material cerámico que actúa como dieléctrico está compuesto de diminutos cristales que se comportan como pequeños dipolos eléctricos (figura 15.2a). Entre los materiales que más se utilizan se encuentran: el sulfato de triglicina (TGS) y el tantanato de litio ($LiTaO_3$), aunque existen otros materiales como los típicos piezoeléctricos: PVF y PVDF. Estos materiales presentan una propiedad interesante: tienen un momento dipolar intrínseco[1] (figura 15.2b) que depende de la temperatura y decrece con esta de una forma no lineal tal como muestra la figura 15.2c. El **efecto piroeléctrico** se produce cuando la radiación incidente hace que se incremente la temperatura del material cerámico, provocando que cambie su polarización interna debido a la agitación térmica. Esta reducción en la polarización ocasiona un decremento de la carga superficial del material y un exceso de carga inducida en los electrodos, Δq. A este mecanismo de generación de carga se le denomina piroelectricidad primaria[2].

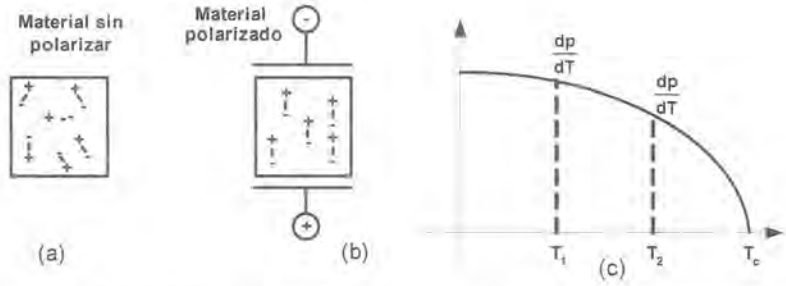


Figura 15.2. (a) Material piroeléctrico sin polarizar; (b) material polarizado y situado entre dos electrodos para formar el sensor; (c) dependencia de la polarización del material piroeléctrico en función de la temperatura. A partir de la temperatura de Curie, T_c desaparece el momento dipolar y el material pierde su polarización de forma permanente [15.3].

El detector no requiere de ninguna fuente de excitación externa, tan sólo se necesita un circuito eléctrico capaz de medir la carga inducida. Sin embargo, la carga sobre los electrodos se va inducir siempre que existan cambios en el momento dipolar y esto ocurre por cambios en la temperatura del material. En condiciones estables, el material piroeléctrico permanece polarizado pero se "descarga" (los portadores libres neutralizan la carga de polarización) en un tiempo relativamente corto, no induciéndose cargas en los electrodos. Por tanto, sólo cuando cambie la temperatura se genera el desequilibrio que permite la generación de carga eléctrica superficial, es decir, se trata de un sensor dinámico.

[1] Este momento dipolar se induce durante la fabricación mediante un proceso similar al que estudió en el capítulo anterior para los materiales piezoeléctricos.

[2] También existe, de forma simultánea, un mecanismo de dilatación térmica que debido al efecto piezoeléctrico induce una carga, se denomina piroelectricidad secundaria.

15.3 Circuito equivalente de un sensor piroeléctrico

Los pasos que tienen lugar para convertir una fuente de radiación en una señal eléctrica, como por ejemplo una tensión, se muestran en la figura 15.3 de forma esquemática:

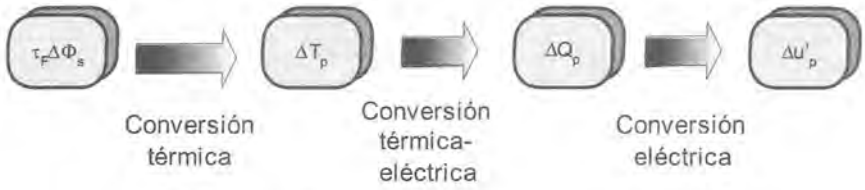


Figura 15.3. Etapas de conversión en un detector piroeléctrico.

El flujo de radiación $\Delta\Phi$ emitido por la fuente atraviesa un filtro infrarrojo de transmisión τ_F (generalmente se selecciona un margen de longitudes de onda sobre las que estamos interesados) y llega al detector que absorbe esta energía y provoca un cambio de temperatura ΔT. Por efecto piroeléctrico, se produce la conversión térmica-eléctrica generándose una carga en los electrodos del sensor que posteriormente se transforma en una señal de tensión.

Modelo térmico

En la figura 15.4a se muestra el sensor piroeléctrico con sus parámetros térmicos fundamentales [15.1]. El sensor se caracteriza por un coeficiente de absorción α, una capacidad calorífica H_p, un grosor d, un área A_s, una conductancia térmica δ y con T_A (temperatura ambiente). El circuito térmico equivalente del sensor se muestra en la figura 15.4b.

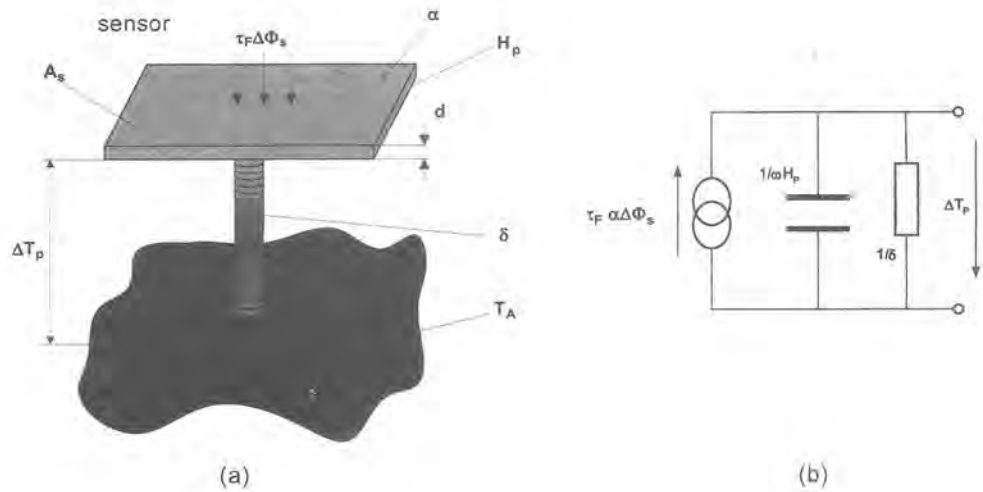


Figura 15.4. (a) Sensor piroeléctrico; (b) modelo térmico equivalente.

A partir del modelo expresado en 15.4b se obtiene la constante de tiempo térmica:

$$\tau_T = H_p / \delta \qquad (15.1)$$

y el gradiente de temperatura entre el sensor y la temperatura ambiente:

$$\Delta T_p = \frac{\alpha \tau_F \Phi}{\sqrt{\delta^2 + \omega H_p{}^2}} \tag{15.2}$$

Si se considera una fuente de radiación senoidal Φ_s resulta, en régimen permanente:

$$\Delta T^s{}_p = \frac{\alpha \tau_F \Phi_s}{\delta} \frac{1}{\sqrt{1+(\omega \tau_T)^2}} \tag{15.3}$$

Para obtener la mayor sensibilidad, es decir, para que dado un flujo de excitación se produzca el mayor cambio de temperatura en el sensor (esto induce más carga) deben maximizarse los siguientes parámetros:

1°) $\alpha\tau_F$ debe ser lo mayor posible, es decir, debe tratarse que toda la radiación que llegue, sea absorbida en lo posible por el sensor. Para ello, suele recubrirse el material piroeléctrico con una fina capa negra que incrementa su capacidad de absorción a las longitudes de onda de interés.

2°) Por otro lado, la capacidad calorífica H_p debe ser también baja (así, la constante de tiempo es más pequeña y aumenta la velocidad de respuesta). Esto se consigue reduciendo el grosor d del sensor.

3°) Por último, también debería de reducirse la conductancia térmica o coeficiente de disipación δ, aunque debe llegarse a un compromiso ya que su reducción afecta a la constante de tiempo.

Modelo eléctrico

El proceso de conversión térmica a una magnitud eléctrica se produce por el efecto piroeléctrico y es proporcional al cambio de temperatura, al coeficiente piroeléctrico del material p y a la superficie del detector A, obteniéndose:

$$I_p = \frac{dQ}{dt} = Ap\frac{dT}{dt} \tag{15.4}$$

Suponiendo una excitación senoidal y sustituyendo (15.3) en (15.4) resulta:

$$i_p = \omega pA \frac{\alpha \tau_F \Phi_s}{\delta} \frac{1}{\sqrt{1+(\omega \tau_T)^2}} \tag{15.5}$$

La figura 15.5 representa el modelo eléctrico de un sensor piroeléctrico, donde C_p es la capacidad del sensor y R_G la resistencia de pérdidas que suele ser muy alta (del orden de 10 GΩ).

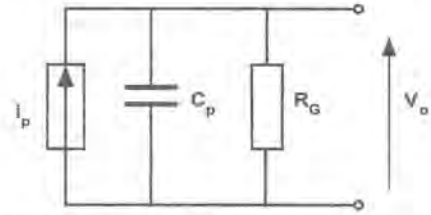


Figura 15.5. Modelo eléctrico del sensor piroeléctrico.

El material con que está construido el sensor depende del margen de la radiación que se pretenda medir. El sulfato de triglicina (TGS) tiene una alta sensibilidad pero una baja temperatura de Curie (49°C); otros materiales piezocerámicos también pueden ser usados como elemento sensor para temperatura desde 100 a

varios cientos de grados centígrados. En la tabla 15.1 se indican los parámetros de algunos de los materiales más comunes.

Material	Coeficiente piroeléctrico nC/cm^2K	Constante dieléctrica	Temperatura de Curie
TGS (Sulfato de triglicina)	$3{,}0 \cdot 10^{-8}$	11	49ºC
TaO_3Li (Tantanato de litio)	$1{,}7 \cdot 10^{-8}$	46	600ºC
PVDF (Polivinilideno)	$27 \cdot 10^{-9}$	12	100ºC

Tabla 15.1. Características principales de algunos materiales piroeléctricos.

15.4 Acondicionamiento de señal

El circuito equivalente del sensor (figura 15.5) se caracteriza por una alta impedancia de salida y una corriente extremadamente baja por lo cual debe utilizarse un amplificador con una alta impedancia de entrada. Por lo general, existen dos alternativas para realizar el acondicionamiento denominadas: modo corriente y modo tensión (figura 15.6). El modo tensión se implementa mediante un seguidor de tensión (figura 15.6a) y el modo corriente mediante un convertidor corriente-tensión (figura 15.6b).

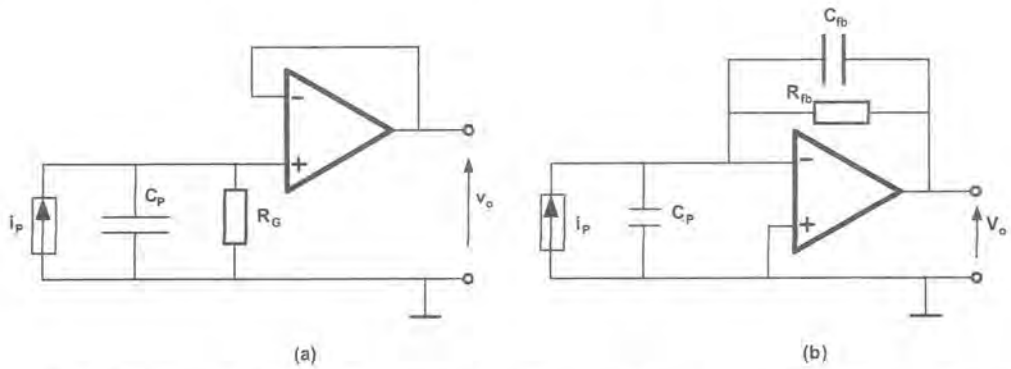


Figura 15.6. Circuitos de acondicionamiento para sensores piroeléctricos, (a) en modo tensión, y (b) en modo corriente.

La señal de salida para ambos modos puede obtenerse a partir de:

$$v_o = \omega\alpha\tau_F\Phi Ap\frac{1}{\delta}R\frac{1}{\sqrt{1+(\omega\tau_T)^2}}\frac{1}{\sqrt{1+(\omega\tau_E)^2}} \tag{15.6}$$

donde τ_E es la constante de tiempo eléctrica. La relación entre la salida y el flujo de radiación se denomina **responsividad** y se expresa como:

$$R_V = \frac{v_o}{\Phi} = \omega\alpha\tau_F Ap\frac{1}{\delta}R\frac{1}{\sqrt{1+(\omega\tau_T)^2}}\frac{1}{\sqrt{1+(\omega\tau_E)^2}} \tag{15.7}$$

donde $R = R_G$ y $\tau_E = R_G C_p$ para el modo tensión y $R = R_{fb}$ y $\tau_E = R_{fb} C_{fp}$ para el modo corriente. En el modo corriente podemos hacer que C_{fb} sea considerablemente más baja que la capacidad del sensor piroeléctrico C_p y, por tanto, la constante de tiempo τ_E (constante de tiempo eléctrica) puede ser más baja que en el modo tensión, es decir, el modo corriente presenta una respuesta más rápida.

En general, las fuentes de ruido del sensor y del amplificador limitan el mínimo flujo de radiación detectable. Como medida de la relación S/N se utiliza habitualmente la detectividad específica[3] D^*. En principio, es irrelevante si se utiliza el modo corriente o el modo tensión en relación con el ruido porque las mismas fuentes están presentes en ambos modos. La única fuente de ruido que siempre está presente, incluso en un sensor ideal, es la que se produce por el intercambio térmico entre el sensor piroeléctrico y su entorno y se denomina ruido de temperatura[4]. Este ruido permite calcular, teóricamente, la mayor detectividad específica de un detector piroeléctrico a temperatura ambiente y resulta $D_{máx}^* = 1{,}8 \cdot 10^{10}$ cmHz$^{1/2}$ W^{-1}.

EJERCICIO 15.1

A un sensor piroeléctrico con C_p = 65 pF, R_G = 15 GΩ y R_V = 300 V/W (10 Hz) se le aplica una excitación de 1 μW (10 Hz). Si se adopta el circuito de la figura 15.6a, ¿qué características debería cumplir el operacional seleccionado? ¿Cuál sería su salida y su constante de tiempo eléctrica?

SOLUCIÓN:

El operacional debería poseer una alta impedancia de entrada, así que deberíamos elegir un operacional con bajas corrientes de polarización (con entrada FET). La salida se calcula directamente a partir de la responsividad de tensión que es un dato que los fabricantes proporcionan habitualmente:

$$v_o = 1\ \mu W \cdot 300\ V/W = 300\ mV$$

La constante de tiempo eléctrica es: $\tau_E = 15\ G\Omega \cdot 65\ pF = 0{,}975$ s.

Consideraciones sobre el modo tensión

El circuito más simple para el modo tensión utiliza un JFET como seguidor de tensión (figura 15.7a). Generalmente, la resistencia de puerta y el FET se integran en el encapsulado del detector, mientras que la resistencia de fuente se sitúa fuera. Este circuito proporciona una alta relación S/N y presenta baja sensibilidad a la temperatura.

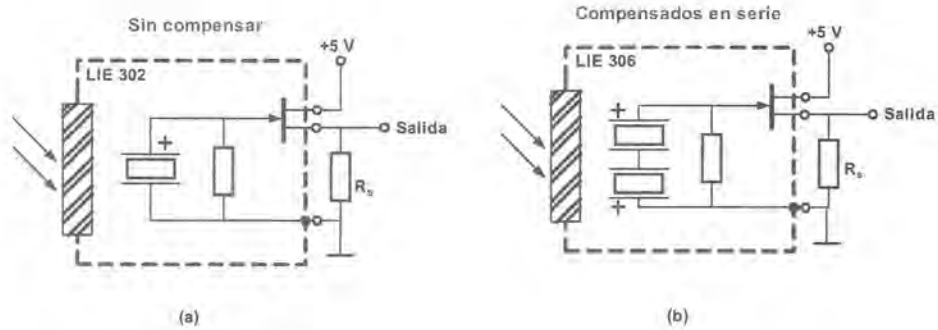


Figura 15.7. Circuitos básicos de sensores piroeléctricos comerciales en modo tensión [15.1].

[3] En [15.1] puede ver una tabla donde se detallan las diferentes fuentes de ruido que intervienen en el cálculo de este parámetro.
[4] $D_T^* = \alpha\tau_E(A/4kT^2\delta)^{1/2}$.

En la figura 15.7b se muestran dos sensores conectados en oposición serie con objeto de cancelar el ruido de modo común. Si la misma magnitud de radiación llega a los dos sensores, generarán idéntica polarización y, debido a su conexión en oposición, la tensión o corriente sobre R se anula; sin embargo, las señales que lleguen a un solo sensor no serán canceladas. La ganancia de estos circuitos se deriva de la transconductancia del JFET (g_m) en el punto de trabajo y de la resistencia de fuente (R_s):

$$A_V \approx \frac{g_m R_s}{1 + g_m R_s} \leq 1$$

Por su simplicidad este circuito (figura 15.7) es el más utilizado en los detectores piroeléctricos. Debe tomarse alguna precaución en relación con la resistencia de carga en la puerta del FET que, en general, debería tener un valor de, al menos, 10 GΩ. Para preservar este valor tan alto debe evitarse que la humedad le afecte y provoque corrientes de pérdidas. Además, estos sensores muestran una gran sensibilidad a interferencias acústicas, térmicas y electromagnéticas por lo que suelen apantallarse o encapsularse herméticamente (sensor y amplificador) para reducir los efectos de los movimientos de aire.

Consideraciones sobre el modo corriente

Los detectores piroeléctricos en modo corriente son menos utilizados que en modo tensión, más que nada porque su circuito es más complejo y la mayoría de las aplicaciones de los sensores piroeléctricos son para la detección de luz o la detección de movimiento. En la figura 15.8 se muestra la configuración típica de un sensor en modo corriente con el elemento sensor y el amplificador incorporado.

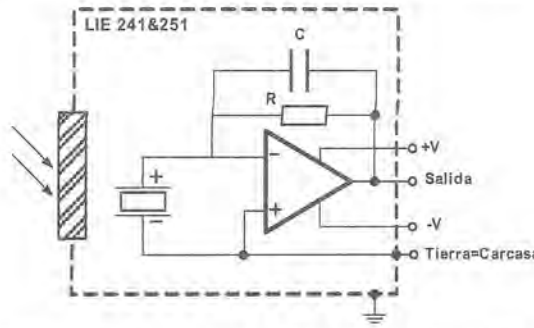


Figura 15.8. Configuración de un sensor piroeléctrico comercial en modo corriente donde se encapsula el sensor y el amplificador [15.1].

En general, el modo corriente presenta mayor ruido que el modo tensión aunque, con los amplificadores operaciones actuales de bajo ruido, puede conseguirse una relación S/N como la del JFET y, además, con mayor responsividad (su constante de tiempo eléctrica puede hacerse pequeña). Las ventajas de este modo pueden resumirse en:

1. Alta responsividad (R_v > 20000 V/W) y también alta estabilidad.
2. Bajo *offset* y bajas derivas térmicas.
3. Baja constante de tiempo.

15.5 Aplicaciones

La aplicación más típica de los sensores piroeléctricos es la detección de radiación infrarroja. Entre las aplicaciones más habituales pueden citarse:

- Pirometría óptica: medida de temperatura a distancia.
- Radiometría: medida de la potencia generada por una fuente de radiación.
- Termometría: termómetros de alta resolución.
- Espectroscopía y análisis de gases (detectores CO_2).
- Tranmisión, reflexión y absorción de IR.
- Detectores de presencia y movimiento.
- Visión nocturna: imágenes térmicas.
- Detectores de llama.

No obstante, los detectores piroeléctricos no se limitan a la detección en el infrarrojo. También tienen aplicaciones en la zona visible y en el ultravioleta (UV), por ejemplo, para controlar la estabilidad de fuentes de radiación [15.5]. También puede utilizarse como detector de partículas con gran sensibilidad y bajo coste [15.7] tanto de átomos neutros como iones (H, H^+, H_2^+, H_3^+ y He^+). Los materiales piroeléctricos también presentan una buena respuesta a bajas temperaturas, así su respuesta a 50 K es del mismo orden de magnitud que a 300 K.

Una de las principales ventajas de los sensores piroeléctricos frente a otros sensores térmicos viene dada por su velocidad de respuesta. Además son bastante simples y su coste es bajo. Sin embargo, y al contrario que en otros sensores térmicos, la salida del sensor piroeléctrico responde a los cambios en la temperatura del material y no al valor estático de la temperatura. La medida estática de temperatura, no obstante, también puede realizarse y se utiliza por ejemplo en los termómetros IR [15.4]. Para ello el flujo de radiación debe pulsarse u obstruirse mecánicamente para generar una serie de pulsos (figura 15.9).

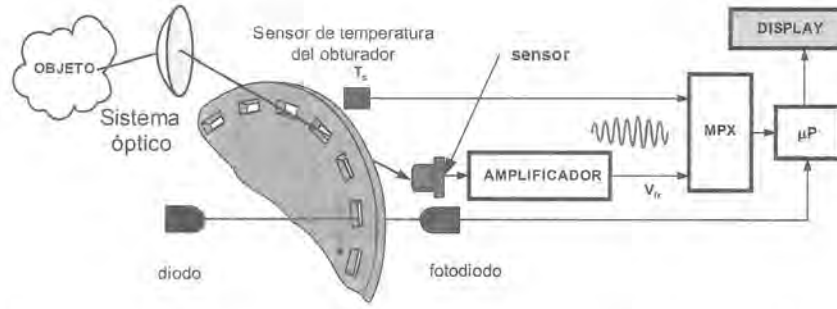


Figura 15.9. Termómetro infrarrojo con obturador giratorio para generar una radiación pulsante.

La salida del amplificador es una señal alterna cuya magnitud depende de la radiación IR incidente y de la velocidad de giro del elemento obstructivo. La rotación de este elemento se sincroniza con un microcontrolador por medio de un optoacoplador y permite convertir la radiación IR en una señal de alterna. En los termómetros denominados instantáneos, muy utilizados para propósitos médicos, se utiliza un procedimiento similar. Se obstruye la radiación al sensor por medio de un diafragma y este no genera salida pero cuando se abre, la radiación llega al sensor y fluye corriente al amplificador en forma de un pulso cuyo valor de pico es proporcional al flujo de la radiación IR recibida en el momento de la apertura del diafragma [15.8]. La temperatura se obtiene así, en un momento determinado y no de forma continua. El principio de funcionamiento de estos termómetros se basa en la ley de Stefan-Boltzmann, que relaciona flujo de radiación con la temperatura. En la figura 15.10 se muestra el circuito de un sensor piroeléctrico en modo corriente útil en aplicaciones de pirometría y fotometría.

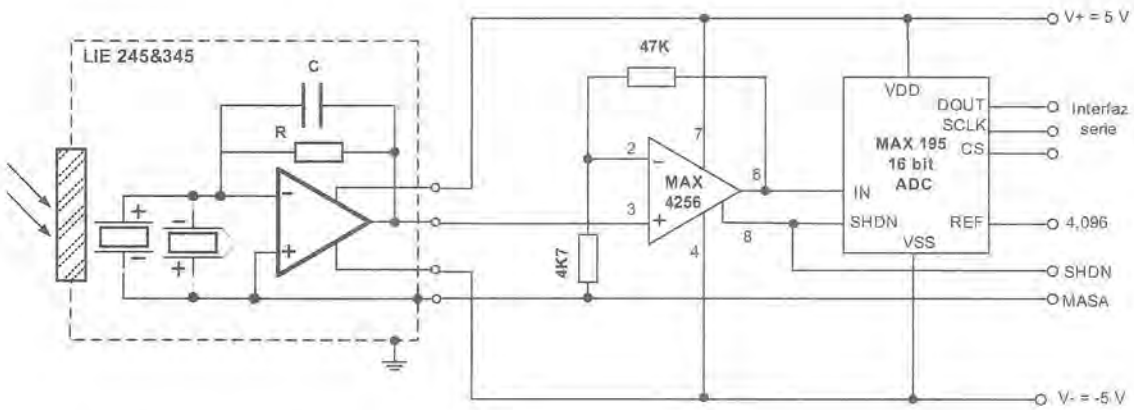


Figura 15.10. Sensor piroeléctrico en modo corriente acoplado directamente a un convertidor A/D [15.1].

Resumen

Los sensores piroeléctricos son sensores térmicos de radiación. Están basados en el efecto piroeléctrico que consiste en la inducción de una carga sobre los electrodos del sensor por efecto térmico. El sensor es un pequeño condensador formando por una pastilla muy fina de material cerámico ferroeléctrico con dos electrodos en su superficie sobre los cuales se induce una carga eléctrica debido al cambio de polarización que experimenta el material ferroeléctrico con la temperatura. Entre los materiales que más se utilizan se encuentra: el sulfato de triglicina (TGS) y el tantalato de litio ($LiTaO_3$).

Entre sus aplicaciones puede citarse la medida de temperatura a distancia, medida de potencia de radiación (radiómetros), detectores de CO_2, detectores de presencia por medio de la detección de radiación infrarroja, termómetros de alta resolución, etc. Comparados con otros sensores termoeléctricos, los sensores piroeléctricos son más simples, más sensibles, su construcción es más robusta y su respuesta más rápida.

El acondicionamiento de señal de salida se realiza de dos formas: en modo corriente y en modo tensión. En el primer caso, se utiliza un convertidor corriente-tensión y su respuesta es más rápida pues no depende de la capacidad del sensor. En el segundo suele utilizarse un seguidor de tensión con JFET. Este último modo es el más utilizado por su simplicidad.

Problemas propuestos

15.1 Indique el tipo de respuesta en frecuencia de la corriente de cortocircuito del sensor piroeléctrico a la vista de la ecuación 15.5.

15.2 En un sensor piroeléctrico el cambio de temperatura frente a la frecuencia es de tipo pasabajo. Calcule la frecuencia de corte si se dispone de los siguientes datos: $\delta = 1{,}98$ mW/K y $H_p = 310$ μWs/K.

15.3 ¿Por qué motivo resulta conveniente reducir el grosor en un sensor piroeléctrico?

15.4 Dibuje un diagrama de Bode de la tensión de salida de un sensor piroeléctrico en la configuración de modo tensión y modo corriente. Datos: $\tau_F = \alpha = 1$; $\Phi_s = 1$ μW; $\delta = 1{,}95$ mW/K; $H_p = 310$ μWs/K; $C_p = 62$ pF; $A_s = 4$ mm^2; d = 25 μm; p = 17 nC/cm^2/K; $R_G = R_{fb} = 24$ GΩ; $\tau_{E\ modo\ tensión} = 1{,}5$ s y $\tau_{E\ modo\ corriente} = 16$ ms.

16

Sensores optoelectrónicos generadores de señal

16.1 Fotodiodos y fototransistores
16.2 Dispositivos de acoplamiento de carga (CCDs)
16.3 Fotomultiplicadores

16.1 Fotodiodos y fototransistores

16.1.1 PRINCIPIO DE FUNCIONAMIENTO DE LOS FOTODIODOS

En la figura 16.1 se muestra la sección transversal de un fotodiodo y el diagrama de bandas correspondiente. Cuando un fotón incide en el fotodiodo con una energía mayor que la correspondiente a la banda prohibida del semiconductor se puede generar un par electrón-hueco; esta generación de pares electrón-hueco ocurre en la zona P, en la N y en la zona de transición. Si ocurre en esta última zona o en sus proximidades, el campo eléctrico acelera los electrones hacia la zona N y los huecos hacia la P, apareciendo una carga negativa en la zona N y positiva en la zona P. Si se conecta un circuito externo entre el cátodo y el ánodo del fotodiodo, los electrones fluirán desde la capa N y los huecos desde la P hasta los electrodos opuestos, dando lugar a una circulación de corriente, también denominada fotocorriente.

Los pares electrón-hueco generados fuera de la zona de transición también pueden alcanzarla por mecanismos de difusión o bien pueden recombinarse antes sin que contribuyan a la fotocorriente. Puesto que el mecanismo de difusión es relativamente lento, los fotodiodos se diseñan para que la probabilidad de recombinación sea elevada con el fin de no comprometer el tiempo de respuesta [16.6].

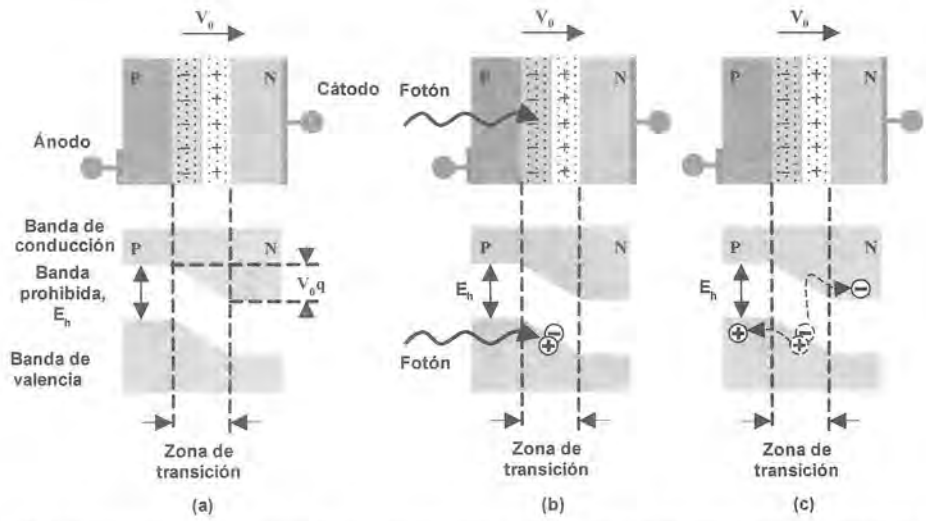


Figura 16.1. (a) Sección transversal de un fotodiodo y diagrama de bandas; (b) generación de un par electrón-hueco al incidir un fotón con energía mayor que la banda prohibida; (c) el electrón es acelerado hacia la zona N y el hueco hacia la zona P debido al campo eléctrico. (V_0 es la tensión de la unión y q es la carga de un electrón, $1{,}6 \cdot 10^{-19}$ C.)

Por otra parte, la necesidad de que la energía del fotón, E_f, deba ser mayor o igual que la de la banda prohibida, E_h,

$$E_f \geq E_h \tag{16.1}$$

supone un límite a la respuesta del fotodiodo. Teniendo en cuenta que la energía de un fotón viene dada por $h\nu$, donde h es la constante de Planck ($6{,}62 \cdot 10^{-34}$ J·s) y ν la frecuencia, se puede demostrar que la longitud de onda de la luz incidente debe de ser menor que un cierto valor que viene determinado por la energía de la banda prohibida del semiconductor de que se trate [16.8]:

$$\lambda(\text{nm}) \leq \frac{1240}{E_h(\text{eV})} \tag{16.2}$$

La longitud de onda máxima se denomina longitud de onda de corte, λ_h:

$$\lambda_h(\text{nm}) = \frac{1240}{E_h(\text{eV})} \tag{16.3}$$

En la tabla 16.1 se muestran los valores de la energía de la banda prohibida de diversos materiales, así como la longitud de onda de corte.

Semiconductor	Energía de la banda prohibida, E_h(eV)	Longitud de onda de corte, λ_h(nm)
Si	1,12	1100
Ge	0,66	1870
InP	1,35	910
InGaAsP	0,89	1400
InGaAs	0,75	1650

Tabla 16.1. Valores de la energía de la banda prohibida y de la longitud de onda de corte para diferentes semiconductores (extraída de [16.9]).

En principio, cabría esperar que para las longitudes de onda menores que la longitud de onda de corte, la respuesta del fotodiodo fuese similar. Sin embargo, ya se ha comentado que el mecanismo esencial por el que se produce la generación de la fotocorriente es la incidencia del fotón en la zona de transición; sin embargo, los fotones con diferentes longitudes de onda no tienen la misma probabilidad de alcanzar esta zona. Así, cuando la luz incide en el fotodiodo, su intensidad va decreciendo exponencialmente a medida que lo atraviesa (figura 16.2a) según la siguiente ecuación:

$$I(x) = I_0 \cdot e^{-\alpha x} \tag{16.4}$$

donde I_0 es la intensidad de la luz incidente, α es el coeficiente de absorción, característico del material de que se trate, y x es la distancia recorrida por la luz desde que incide en la superficie del fotodiodo [16.9]. Al inverso del coeficiente de absorción se le denomina coeficiente de penetración δ,

$$\delta = \frac{1}{\alpha} \tag{16.5}$$

y representa la distancia a la que la intensidad de la luz incidente ha disminuido un 63% del valor inicial I_0, como se puede deducir teniendo en cuenta las expresiones (16.4) y (16.5).

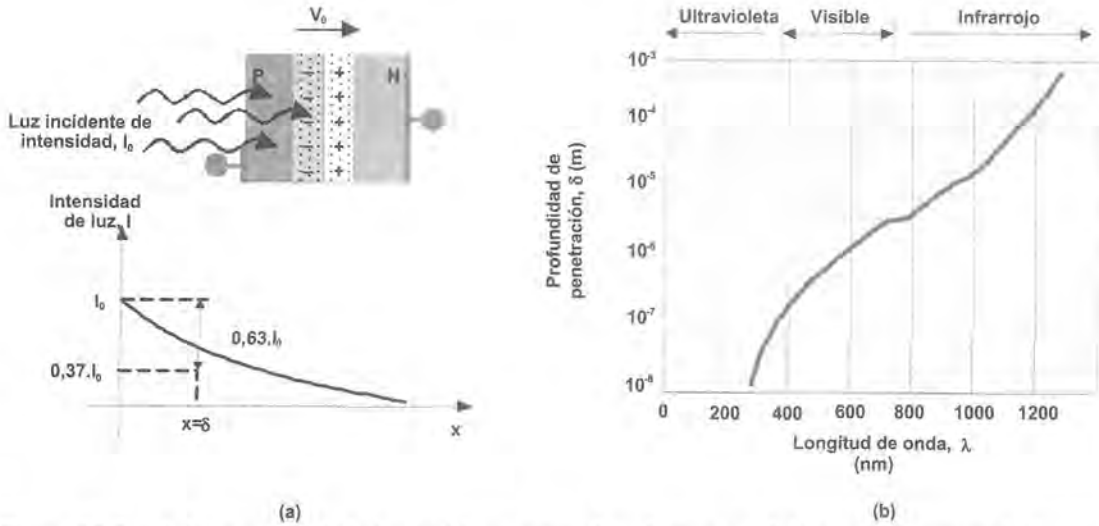


Figura 16.2. (a) Atenuación de la luz incidente en un fotodiodo a medida que penetra en el mismo; (b) profundidad de penetración en el silicio en función de la longitud de onda de la luz incidente (tomado de [16.18]).

La profundidad de penetración depende del material y de la longitud de onda de la luz. En la figura 16.2b se muestra su valor para el silicio y se puede observar que la luz de una longitud de onda pequeña (como la luz ultravioleta) se absorbe antes, mientras que las longitudes de onda más largas (como la luz infrarroja) penetran más en el cristal. Teniendo en cuenta que la fotocorriente se debe, en su mayor parte, a la incidencia de la luz en la zona de transición, para que un fotodiodo tenga una buena respuesta a longitudes de onda cortas la zona de transición puede ser corta y debe estar situada muy próxima a la superficie del fotodiodo; mientras que la zona de transición debe ser larga si se desea una buena respuesta para longitudes de onda largas.

Mientras que las características de los fotodiodos mencionadas hasta el momento condicionan su respuesta estática frente a la luz, un parámetro de gran importancia y que determina en buena medida la respuesta dinámica es la capacidad de transición. En la figura 16.3a puede observarse la estructura de un fotodiodo con la correspondiente zona de transición, la distribución de la carga y la tensión de la unión, V_o. Como se puede apreciar en la misma figura, esta distribución es idéntica a la de un condensador de placas planas separadas una distancia L_{zt}, donde L_{zt} es la longitud de la zona de transición.

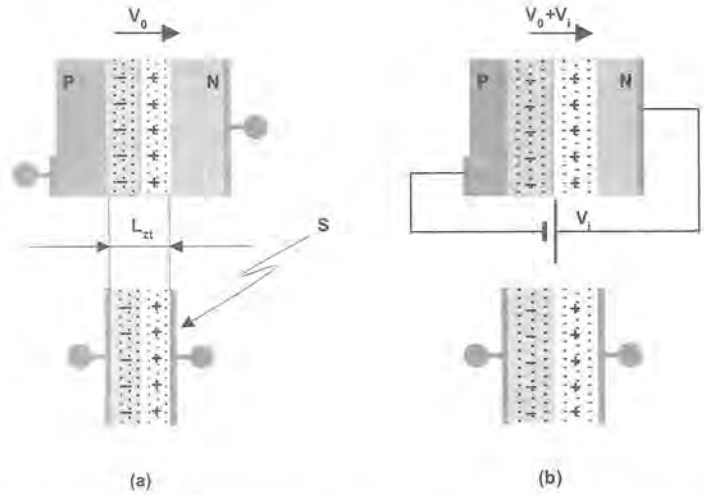


Figura 16.3. (a) Símil entre la distribución de carga en la zona de transición de un fotodiodo y un condensador de placas planas; (b) al aplicar tensión inversa aumenta la longitud de la zona de transición lo que equivale a separar la placas del condensador.

Se puede demostrar que la capacidad de transición del fotodiodo, C_t, en función de la tensión inversa aplicada, V_i, es [16.6]:

$$C_t = C_0 \frac{1}{\sqrt{1+V_i/V_0}} \tag{16.6}$$

donde C_o es la capacidad de transición cuando no se aplica tensión al fotodiodo. La ecuación (16.6) tiene un importante significado puesto que indica que la capacidad de transición disminuye al aplicar tensión inversa.

16.1.2 TIPOS DE FOTODIODOS

Existen, fundamentalmente, cuatro tipos de fotodiodos (tabla 16.2): el fotodiodo PN, el PIN, el Schottky y el de avalancha [16.8]. A continuación, se resumen las características más destacables de cada uno de ellos:

Tipo	Estructura	Características
PN		• Buenas características generales
PIN		• Tensión inversa 5~100 V • Capacidad de transición pequeña • Muy rápidos
Schottky		• Buena respuesta en el ultravioleta
Avalancha		• Tensión inversa 150~200 V • Ganancia ~ 100 • Muy rápidos

Tabla 16.2. Tipos de fotodiodos, estructura y características destacables.

- Fotodido PN: la estructura corresponde a la estudiada en el apartado anterior. Se trata de un fotodiodo de unas buenas características generales y el más utilizado dentro del campo de la instrumentación electrónica.
- Fotodiodo PIN: en este fotodiodo existe una capa de semiconductor intrínseco entre la capa P y la N. La zona de transición se extiende a lo largo de toda esta capa para lo que, normalmente, es necesario aplicar una cierta tensión inversa (entre 5 y 100 V dependiendo del fotodiodo). Puesto que la zona de transición es larga, la respuesta a longitudes de onda largas (luz infrarroja) es muy buena y, además, la capacidad de transición es pequeña por lo que son fotodiodos muy rápidos. Son muy utilizados en comunicaciones.
- Fotodiodo Schottky: en este fotodiodo la zona de transición está, por la propia estructura del diodo, muy próxima a la superficie por lo que la respuesta a las longitudes de onda cortas (como las correspondientes a la luz ultravioleta) es buena.
- Fotodiodo de avalancha: si a un fotodiodo se le aplica una tensión inversa suficientemente elevada, los electrones generados por la incidencia de luz en la zona de transición son fuertemente acelerados debido al campo eléctrico existente en la zona y pueden alcanzar una energía tal que den lugar a nuevos pares electrón-hueco al chocar con otros átomos del semiconductor. Por el mismo proceso, estos nuevos pares pueden dar lugar a otros y así sucesivamente. Es decir, se produce un fenómeno multiplicativo consiguiéndose una ganancia determinada. Por ejemplo, para fotodiodos de silicio la ganancia suele rondar el valor 100.

  Los fotodiodos diseñados para trabajar en zona de avalancha suelen tener una estructura $NP\pi P^{+}$, donde la capa π es de tipo P pero ligerísimamente dopada. Un inconveniente de importancia reside en que la ganancia en la zona de avalancha depende mucho de la temperatura por lo que en muchos casos es necesario refrigerar el fotodiodo y mantener la temperatura constante o bien regular la tensión inversa aplicada en función de la temperatura para mantener la ganancia.

16.1.3 MODELO ELÉCTRICO DE UN FOTODIODO

Para modelar el fotodiodo hay que tener en cuenta, entre otros, los siguientes aspectos [16.2], [16.8], [16.18], (figura 16.4):

1. El comportamiento del diodo como tal, cuya curva característica corriente-tensión (i_d-v_d) viene dada por la siguiente ecuación:

$$i_d = I_s\left(e^{\frac{qv_d}{kT}} - 1\right) \tag{16.7}$$

   donde I_s es la corriente inversa de saturación, q es la carga del electrón, k la constante de Boltzmann[1] y T la temperatura absoluta. Esta curva corresponde al comportamiento del fotodiodo en oscuridad.

2. La fotocorriente creada por la incidencia de la luz, que se puede modelar como una fuente de corriente en paralelo con el diodo.

3. La capacidad de transición, que se puede modelar por un condensador en paralelo con los dos elementos anteriores.

4. Además, el comportamiento en zona inversa del fotodiodo tiene un gran interés puesto que es la zona más habitual de trabajo como se verá en el apartado dedicado al acondicionamiento de señal. Para un mejor modelado de esta zona, se puede considerar una resistencia R_p, en paralelo con el fotodiodo, que refleje el incremento de corriente de oscuridad con el incremento de tensión inversa debido a fugas y que, además, permitirá evaluar el ruido térmico en el fotodiodo.

5. La resistencia serie, R_s, debida al semiconductor y a los contactos.

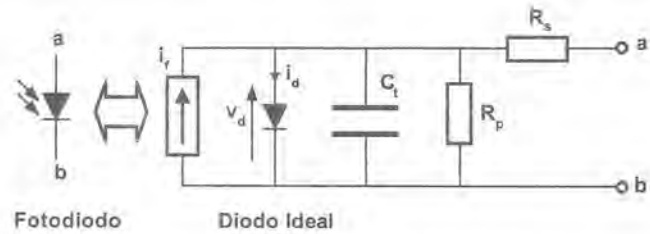


Figura 16.4. Modelo de un fotodiodo.

Las curvas características corriente-tensión típicas de un fotodiodo se pueden observar en la figura 16.5.

Simplificaciones en zona directa

En la práctica, la resistencia en paralelo tiene un valor tan elevado (del orden de decenas a miles de MΩ) y la resistencia serie tiene un valor tan reducido (del orden de una a varias decenas de ohmios) que ambas se suelen despreciar para realizar los análisis (figura 16.6). Sin embargo, no se puede prescindir en zona directa del comportamiento como diodo ni, por supuesto, de la fotocorriente puesto que este último aspecto es el más característico del fotodiodo. Además, la capacidad del fotodiodo no suele ser despreciable salvo que se midan señales de luz continuas.

[1] $k = 1{,}38 \cdot 10^{-23}\ J \cdot K^{-1}$.

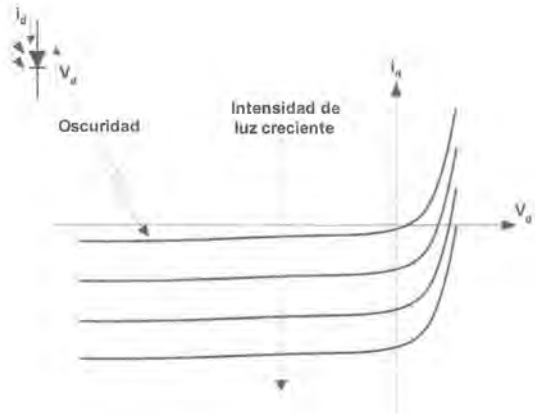


Figura 16.5. Curvas características corriente-tensión de un fotodiodo.

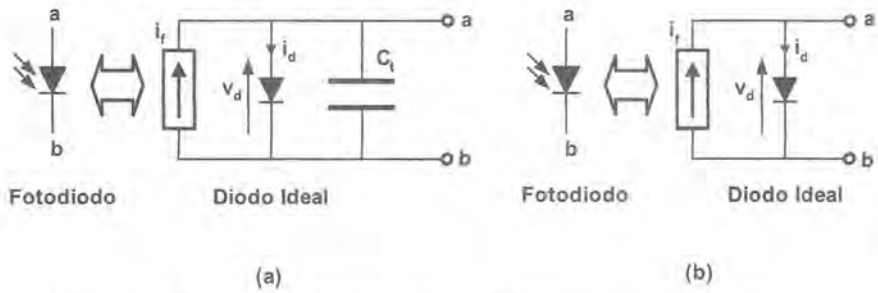


Figura 16.6. Modelos de fotodiodo más habituales en zona directa.

Simplificaciones en zona inversa

Cuando se trabaja en zona inversa también son despreciables, en general, la resistencia paralelo y serie. Además, se suele considerar que el comportamiento como diodo es equivalente a un circuito abierto o, a lo sumo, una fuente de corriente cuyo valor es la corriente inversa de saturación. Los modelos más habituales son, por tanto, los que se muestran en la figura 16.7.

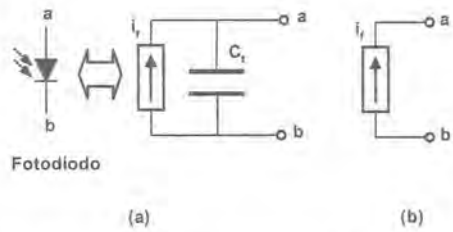


Figura 16.7. Modelos de fotodiodo más habituales en zona inversa.

16.1.4 CARACTERÍSTICAS DE LOS FOTODIODOS

A continuación, se describen las principales características de los fotodiodos que figuran en los catálogos y se indican los valores habituales.

- Superficie activa (*active area*): la superficie activa es el área del semiconductor del fotodiodo expuesta a la luz. Los valores habituales van desde 0,1 mm^2 hasta 100 mm^2. Una superficie grande facilita la recogida de la luz y puede evitar o simplificar la óptica para enfocar la luz incidente en el fotodiodo. Sin embargo, la mayor parte de las características del fotodiodo empeoran con la superficie por lo que conviene ajustarse al valor más pequeño necesario.

- Eficiencia cuántica (*quantum efficiency*): la eficiencia cuántica, η, se define como la relación entre el número de pares electrón-hueco efectivos (es decir, que contribuyen a la fotocorriente) y el de fotones incidentes:

$$\eta = \frac{\text{Nº de pares electrón-hueco efectivos}}{\text{Nº de fotones incidentes}} \tag{16.8}$$

También se puede expresar como la relación entre el número de electrones efectivos por segundo y el número de fotones incidentes por segundo con la ventaja de que estas magnitudes son fáciles de medir [16.9]. Así, el número de electrones efectivos por segundo es la corriente I_f, dividida entre la carga de un electrón q; mientras que el número de fotones incidentes es la potencia de la luz P_0, dividida entre la energía de un fotón hν, es decir:

$$\eta = \frac{\text{Nº de electrones efectivos por segundo}}{\text{Nº de fotones incidentes por segundo}} = \frac{I_f/q}{P_0/h\nu} \tag{16.9}$$

La eficiencia cuántica siempre es menor que uno ya que no todos los fotones incidentes dan lugar a un par electrón-hueco efectivo. En un fotodiodo de silicio, suele rondar el 90-95% para longitudes de onda entre 700 y 900 nm.

- Sensibilidad (*photo sensitivity*): un parámetro directamente relacionado con la eficiencia cuántica, aunque más cómodo de utilizar para el diseñador, es la sensibilidad S, que se define como la relación entre la fotocorriente generada y la potencia de la luz incidente:

$$S = \frac{\text{Fotocorriente (A)}}{\text{Potencia de la luz incidente (W)}} = \frac{I_f}{P_0} \tag{16.10}$$

Teniendo en cuenta las expresiones (16.9) y (16.10) se obtiene:

$$S = \eta\frac{q\lambda}{hc} = \eta\frac{\lambda(\text{nm})}{1240} \tag{16.11}$$

expresión muy interesante y que indica que la sensibilidad aumenta con la longitud de onda λ (aunque siempre con el límite de la longitud de onda de corte λ_h).

A la representación de la sensibilidad en función de la longitud de onda se le denomina la respuesta espectral (*spectral response*) y es una de las características más significativas para seleccionar un fotodiodo. En la figura 16.8 se muestra la respuesta espectral de un fotodiodo de silicio comercial, la máxima sensibilidad teórica según la expresión (16.11) y el límite que impone la longitud de onda de corte, λ_h. Como se puede observar, la eficiencia a longitudes de ondas cortas es más pequeña que a longitudes de onda largas como ya se había previsto en el apartado anterior.

- Capacidad (*terminal capacitance*): la capacidad C del fotodiodo es la suma de la capacidad de transición más la capacidad parásita entre los terminales debida al encapsulado, aunque es la primera la que determina la capacidad total. Se trata de un término muy importante para determinar la velocidad de respuesta. Como ya se ha mencionado, disminuye con la tensión inversa aplicada y aumenta con la superficie del fotodiodo. Los valores típicos van desde algún pF hasta nF.

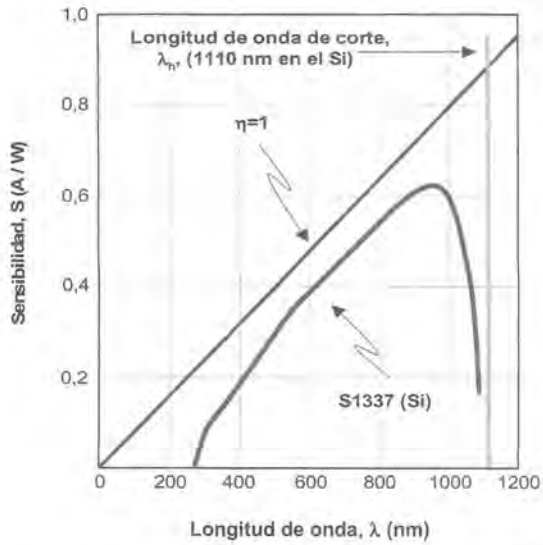


Figura 16.8. Representación de la respuesta espectral para un fotodiodo ideal con una eficiencia cuántica igual a uno ($\eta = 1$), del límite impuesto por la longitud de onda de corte para un fotodiodo de silicio y respuesta real del fotodiodo S1337 de Hamamatsu [16.8].

- Tiempo de subida (*rise time*): se define como el tiempo necesario para que la fotocorriente medida en el circuito externo al fotodiodo pase del 10% al 90% del valor final ante un escalón de luz incidente (figura 16.9).

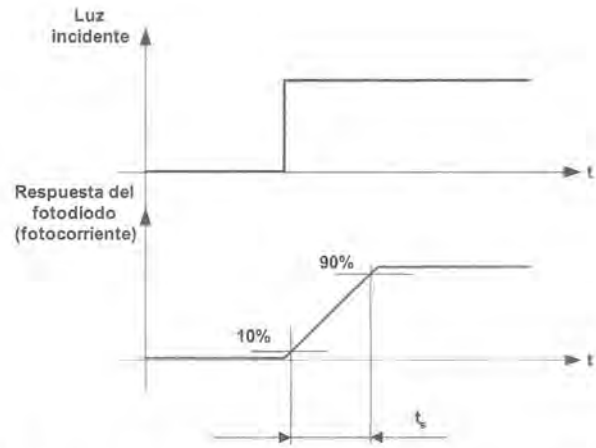


Figura 16.9. Respuesta de un fotodiodo ante un escalón de luz incidente.

El tiempo de subida viene determinado por los siguientes factores [16.8], [16.18]:

1. El tiempo t_t, que tardan los portadores generados en la zona de transición en alcanzar los terminales del fotodiodo.
2. El tiempo de difusión t_d, de los portadores generados fuera de la zona de transición. Evidentemente, este tiempo es mucho mayor que el anterior.

3. El tiempo t_{RC}, determinado por el circuito RC constituido por la capacidad del fotodiodo C, y la resistencia de carga R_c, del circuito externo ($t_{RC} = 2{,}2R_cC$) (figura 16.10).

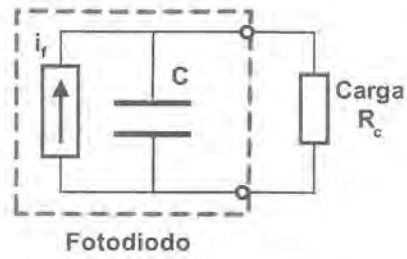


Figura 16.10. Circuito equivalente cuando se conecta una resistencia de carga al fotodiodo, considerando la capacidad de transición.

Los fabricantes de fotodiodos indican en sus catálogos el tiempo de subida para una resistencia de carga determinada, normalmente 1 kΩ. Los valores están comprendidos, generalmente, entre 0,1 μs y decenas de microsegundos para fotodiodos de tipo PN.

- Corriente de oscuridad (*dark current*): la corriente de oscuridad es la corriente que circula por el fotodiodo al aplicar una tensión inversa sin que incida luz. Aumenta con la tensión inversa aplicada y con la superficie del fotodiodo. Además, depende mucho de la temperatura y, aproximadamente, dobla su valor por cada 10°C trabajando en zona inversa. Los valores típicos suelen estar comprendidos entre varios pA y algunas decenas de nA a 25°C, dependiendo de las características del fotodiodo.
- Ruido (*noise*): el ruido en los fotodiodos viene dado por la suma del ruido térmico o ruido Johnson i_{rj}, y el ruido de Schottky i_{rs}, (*shot noise*). El primero se puede calcular a partir de la resistencia paralelo:

$$i_{rj} = \sqrt{\frac{4kT \cdot B_r}{R_p}} \tag{16.12}$$

donde k es la constante de Boltzmann, T la temperatura en Kelvin y B_r el ancho de banda equivalente para el ruido. El ruido Schottky se puede calcular a partir de la corriente que circula por el dispositivo I_D, incluyendo la corriente de oscuridad:

$$i_{rs} = \sqrt{2qI_D B} \tag{16.13}$$

El valor eficaz de la corriente de ruido total, i_{rt}, es:

$$i_{rt} = \sqrt{i_{rj}^2 + i_{rs}^2} \tag{16.14}$$

En los fotodiodos de avalancha es necesario tener también en cuenta el ruido de avalancha [16.8], [16.14].

No obstante, los fabricantes de fotodiodos suelen indicar en sus hojas de características la Potencia Equivalente de Ruido (*Noise Equivalent Power, NEP*) que se define como la relación entre la corriente de ruido i_{rt}, expresada en $A/Hz^{1/2}$ y la sensibilidad S, expresada en A/W. Por lo tanto, vendrá expresada en $W/Hz^{1/2}$:

$$NEP = \frac{i_{rt}}{S} \ W/\sqrt{Hz} \tag{16.15}$$

Los valores típicos están comprendidos entre $10^{-11}W/Hz^{1/2}$ y $10^{-15}W/Hz^{1/2}$. Aumenta con la superficie del fotodiodo y con la tensión inversa puesto que aumenta el ruido Schottky.

EJERCICIO 16.1

Si en un fotodiodo la potencia equivalente de ruido es de 10^{-12} $W/Hz^{1/2}$ y el ancho de banda equivalente del ruido el circuito de medida es de 10^4 Hz, ¿cuál es el mínimo nivel de luz que se puede medir?

SOLUCIÓN:

El ruido equivale a $10^{-12}\cdot(10^4)^{1/2} = 100$ pW de luz incidente; por lo tanto, el mínimo nivel de luz que se puede medir es de 100 pW (relación señal/ruido = 1) sin recurrir a técnicas de tratamiento de señal para disminuir el ruido (por ejemplo, el promediado de señales).

16.1.5 ACONDICIONAMIENTO DE FOTODIODOS

16.1.5.1 Generalidades

La medida de la luz incidente con un fotodiodo se puede realizar de múltiples formas. Una de las más habituales, por las ventajas que presenta, es la medida de la corriente que circula por el fotodiodo sometido a tensión nula (en cortocircuito) (figura 16.11a). La recta de carga correspondiente se muestra en la figura 16.11c.

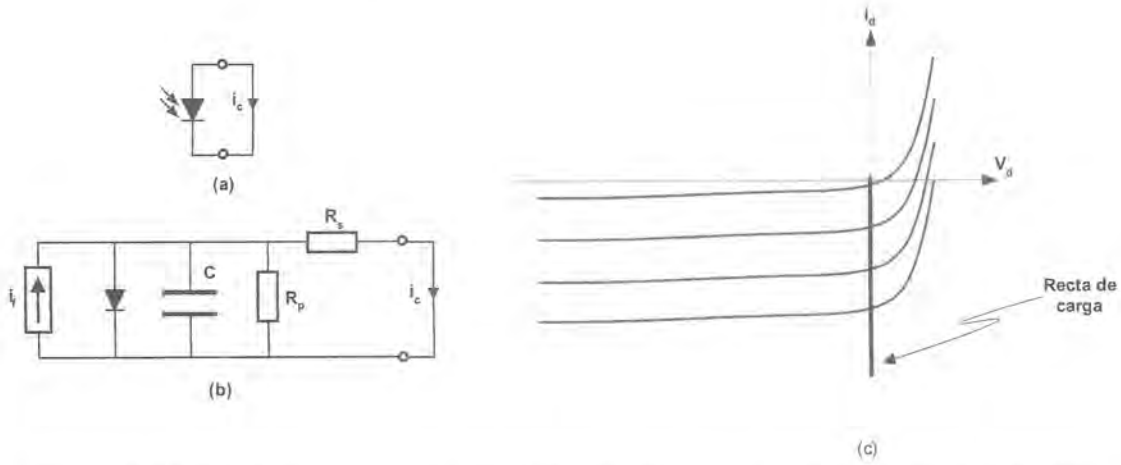


Figura 16.11. Fotodiodo en cortocircuito. (a) Circuito; (b) circuito equivalente según el modelo del fotodiodo; (c) recta de carga.

Teniendo en cuenta el modelo del fotodiodo (figura 16.11b) la corriente de cortocircuito i_c es:

$$i_c = i_f - I_s\left(e^{\frac{qi_cR_s}{kT}} - 1\right) - \frac{i_cR_s}{R_p} \qquad (16.16)$$

En esta ecuación, el segundo término es aproximadamente cero y también lo es el tercer término (debido a los valores habituales de la resistencia serie R_s, y de la resistencia paralelo R_p). Por tanto, la corriente medida es aproximadamente igual a la fotocorriente:

$$i_c \approx i_f \qquad (16.17)$$

Por otra parte, la fotocorriente es lineal con la luz incidente y, aunque los términos despreciados limitan la linealidad, esta es, de todos modos, tremendamente elevada, siendo el error por este concepto menor del 1% en un margen de ocho o diez órdenes de magnitud entre 1 pW y 10 mW [16.8].

Como factores positivos adicionales se tienen los siguientes.

- La dependencia de la temperatura es muy reducida puesto que los únicos términos de la ecuación 16.16 que dependen de la temperatura son los que se han despreciado (la fotocorriente no varía con la temperatura).
- La corriente de oscuridad es pequeña puesto que no se aplica tensión inversa y el ruido también es relativamente pequeño por el mismo motivo.

El mayor inconveniente del circuito analizado reside en que la capacidad de transición es relativamente elevada lo que limita el tiempo de respuesta. Se puede reducir la capacidad aplicando tensión inversa al fotodiodo aunque aparecen otros inconvenientes ligados a una mayor corriente de oscuridad, ruido, etc. Esta y otras posibilidades se muestran en la tabla 16.3 y se comparan entre sí a partir de los resultados obtenidos mediante un análisis similar al llevado a cabo para el circuito anterior.

A la vista de la tabla 16.3, se puede concluir que los circuitos de acondicionamiento más interesantes se basan en la medida de la corriente a tensión cero o bien la corriente cuando el fotodiodo soporta una determinada tensión inversa. La medida en zona directa del fotodiodo resulta, en general, desaconsejable debido a la falta de linealidad y a que la capacidad es elevada; sin embargo, el ruido es muy reducido y en ocasiones resulta la zona más adecuada para medir niveles muy bajos de luz.

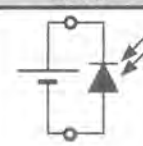
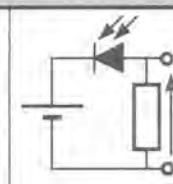
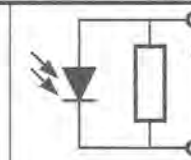
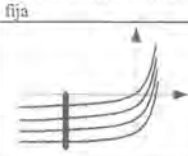
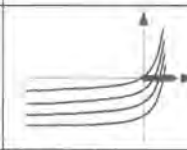
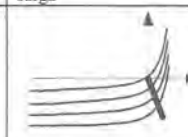
	Zona de trabajo				
	Inversa			Directa	
Circuito	(a)	 (b)	 (c)	(d)	 (e)
Objetivo	Medida de la corriente a tensión cero	Medida de la corriente con una tensión inversa fija	Medida de la corriente en zona inversa sobre una resistencia de carga	Medida de la tensión en circuito abierto	Medida en zona directa sobre una resistencia de carga
Recta de carga					
Linealidad	Excelente	Buena	Aceptable	No es lineal	Media-Baja
Dependencia con la temperatura	Muy baja	Baja	Baja	Muy elevada	Elevada
Corriente de oscuridad	Baja	Elevada	Media	Muy baja (*)	Muy baja (*)
Tiempo de respuesta	Medio	Muy pequeño	Medio	Alto	Alto
Ruido	Medio	Alto	Alto	Bajo	Bajo
Comentarios	Excelentes características generales.	Adecuado cuando se requiere un ancho de banda elevado.	Muy simple. Aplicaciones poco exigentes.	No es lineal. Poco recomendable.	Comportamiento intermedio entre los circuitos (a) y (d). Ruido reducido.

Tabla 16.3. Comparación entre diferentes alternativas para efectuar la medición con fotodiodos. (*) En la medida de la tensión en el fotodiodo en zona directa, el efecto de la corriente de oscuridad se pone de manifiesto en la tensión que aparece en el fotodiodo. En estos casos, sería más conveniente hablar de tensión de oscuridad.

16.1.5.2 Amplificador corriente-tensión (amplificador de transimpedancia)

En el apartado anterior se ha concluido que los circuitos de acondicionamiento más interesantes para efectuar la medición de luz con un fotodiodo se basan en la medida de la corriente que circula por el mismo a tensión cero o bien soportando una determinada tensión inversa. Los circuitos utilizados para efectuar la conversión de la corriente a tensión en estas condiciones se muestran en la figura 16.12. A estos circuitos se les suele denominar amplificadores corriente-tensión o amplificadores de transimpedancia[2]. El análisis de ambos circuitos es similar por lo que el estudio se centrará en el análisis del circuito de la figura 16.12a.

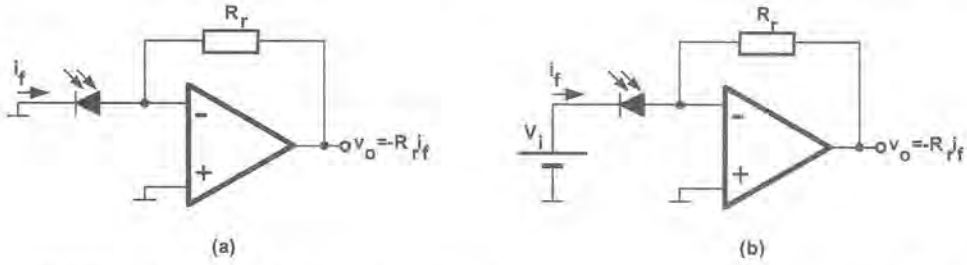


Figura 16.12. (a) Estructura básica del circuito de acondicionamiento comúnmente utilizado para medir la fotocorriente generada en un fotodiodo a tensión cero, denominado amplificador corriente-tensión o amplificador de transimpedancia. Observe que los terminales del fotodiodo se pueden invertir para que la tensión de salida, V_o, sea positiva; (b) circuito de acondicionamiento para medir la fotocorriente con el fotodiodo sometido a una tensión inversa V_i. En este caso, la inversión de los terminales del fotodiodo obligaría a aplicar una tensión inversa V_i, negativa.

En el circuito de la figura 16.12a, el fotodiodo está sometido a una tensión virtualmente cero. Suponiendo que el amplificador operacional sea ideal, la tensión de salida es igual al producto de la resistencia de realimentación R_r, por la corriente que circula por el fotodiodo que es aproximadamente igual a la fotocorriente i_f (a la que también se añade la corriente de oscuridad y la corriente de ruido).

La fotocorriente es, generalmente, muy pequeña (del orden de nanoamperios a microamperios), por lo que la resistencia de realimentación necesaria suele ser muy elevada siendo habituales valores desde centenas de kiloohmios hasta decenas, o incluso, centenas de megaohmios. Por lo tanto, el error de continua debido a las corrientes de polarización del operacional puede ser significativo, lo que obliga, en general, a utilizar amplificadores de entrada FET.

Para disminuir el error debido a las corrientes de polarización, puede recurrirse a situar una resistencia en la entrada no inversora del operacional del mismo valor que la resistencia de realimentación R_r, de modo que el error se debe únicamente a la corriente de desviación. Sin embargo, esta posibilidad presenta algunos inconvenientes [16.6]:

- Un valor de R_r elevado supone que el ruido interno de la propia resistencia sea también elevado y, además, la corriente de ruido de la entrada no inversora del operacional se ve muy amplificada. En caso de recurrir a la compensación, se debería situar un condensador de buenas características en paralelo con la resistencia de compensación para limitar el ancho de banda para el ruido en la entrada no inversora (figura 16.13).

[2] La denominación de amplificador de transimpedancia que se aplica al amplificador corriente-tensión puede resultar, en ocasiones, confusa. Los amplificadores operacionales realimentados en corriente (CFA, *Current Feedback Amplifier*) también se denominan amplificadores de transimpedancia. Sin embargo, el amplificador operacional utilizado en los circuitos de acondicionamiento para fotodiodos es, normalmente, realimentado en tensión (VFA, *Voltage Feedback Amplifier*).

- La caída de tensión en la resistencia de compensación debido a la corriente de polarización del operacional provoca una pequeña caída de tensión en el fotodiodo con lo que la zona de funcionamiento se desplaza ligeramente respecto a la pretendida.

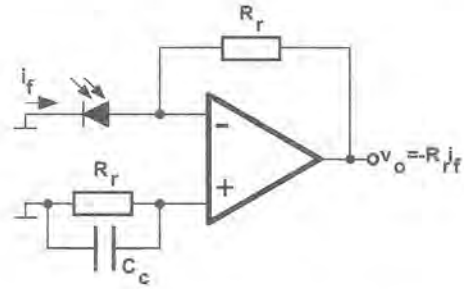


Figura 16.13. *Amplificador corriente-tensión con una resistencia R_r en la entrada no inversora del operacional para disminuir los errores originados por las corrientes de polarización y un condensador C_c, para limitar el incremento en el ruido interno ocasionado por la presencia de esta resistencia.*

En cualquier caso, la compensación no afecta al análisis del circuito salvo en las cuestiones comentadas por lo que, en lo siguiente, se continuará el análisis sin efectuar la compensación.

Por otra parte, los circuitos estudiados presentan una realimentación, β, igual a uno (el fotodiodo se comporta como una fuente de corriente) por lo que, en principio, el ancho de banda es igual al ancho de banda a ganancia unitaria B de un operacional compensado.

Si se efectúa el montaje de estos circuitos, lo más probable sería que se observase un ruido anormalmente elevado o incluso que oscilase, lo que lleva a concluir que el análisis efectuado ha sido excesivamente simplificado. La razón de tal desviación sobre lo previsto teóricamente reside en que no se ha tenido en cuenta la capacidad del fotodiodo que, como se demostrará, puede llevar a que el circuito sea inestable. Considerando esta capacidad, el circuito a analizar se muestra en la figura 16.14.

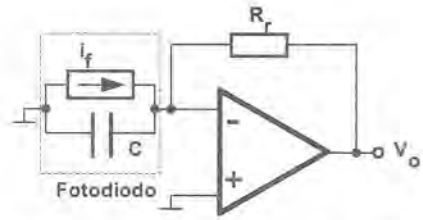


Figura 16.14. Esquema del amplificador corriente-tensión en el que el fotodiodo se ha modelado como una fuente de corriente, i_f, y la capacidad, C, en paralelo.

La realimentación, β, viene ahora dada por:

$$\beta = \frac{X_C}{X_C + R_r} \tag{16.18}$$

donde X_C es la impedancia del condensador. Para determinar la respuesta en frecuencia del operacional interesa conocer la relación $1/\beta$ que es igual a:

$$1/\beta = 1 + R_r C s \tag{16.19}$$

y que presenta un cero a la frecuencia, f_1, dada por:

$$f_1 = \frac{1}{2\pi R_r C} \tag{16.20}$$

Si se representa la curva $1/\beta$ sobre la curva de ganancia en bucle abierto del operacional $|A_{BA}|$ (figura 16.15) se puede observar que ambas curvas se cortan con una pendiente de 40 dB/década lo que supone un margen de fase menor de 45°, por lo que el circuito puede resultar inestable.

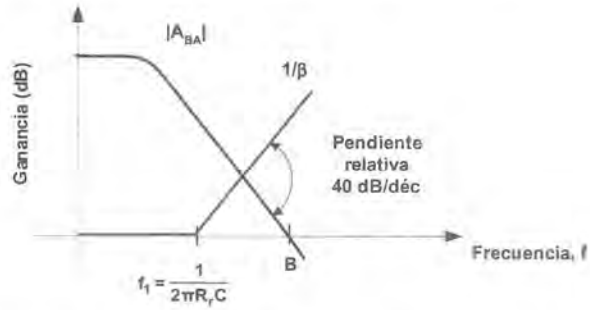


Figura 16.15. Ganancia en bucle abierto del amplificador operacional $|A_{BA}|$, del convertidor corriente-tensión y de la realimentación $1/\beta$, considerando la capacidad del fotodiodo. Como se puede observar, la pendiente relativa entre ambas curvas en el píxel de corte es mayor de 20 dB/déc por lo que el sistema puede ser inestable.

Para estabilizar el sistema, sería necesario añadir un polo a la curva $1/\beta$ antes de su corte con la curva de ganancia del operacional. Para ello basta con añadir un condensador C_r, del valor adecuado, en paralelo con la resistencia de realimentación, R_r [16.2], [16.6], [16.8] (figura 16.16).

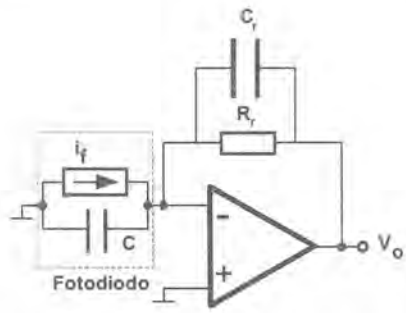


Figura 16.16. Amplificador corriente-tensión con compensación de fase añadiendo un cero en la realimentación mediante el condensador C_r, para compensar el polo añadido por la capacidad del fotodiodo C.

Analizando el circuito de la figura 16.16 se puede demostrar fácilmente que la curva $1/\beta$ viene dada por:

$$1/\beta = \frac{1 + R_r (C + C_r)s}{1 + R_r C_r s} \tag{16.21}$$

que presenta un cero a la frecuencia f_1:

$$f_1 = \frac{1}{2\pi R_r (C + C_r)} \tag{16.22}$$

y un polo a la frecuencia f_2 ($f_2 > f_1$) igual a:

$$f_2 = \frac{1}{2\pi R_r C_r} \tag{16.23}$$

por lo que resulta posible efectuar la compensación pretendida. El problema reside en calcular el condensador C_r de forma que el polo introducido en la curva $1/\beta$ esté situado antes del corte de esta curva con la de la ganancia en bucle abierto del operacional. Para hacerlo, es necesario tener en cuenta que el límite de estabilidad (para margen de fase de 45°) corresponde a la situación que se muestra en la figura 16.17 en la que la curva $1/\beta$ corta a la curva de ganancia en bucle abierto del operacional justo a la frecuencia del polo f_2.

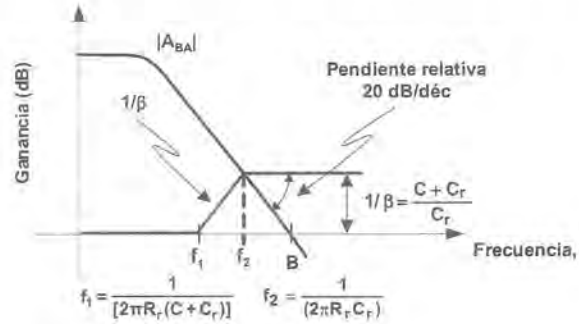


Figura 16.17. Ganancia en bucle abierto del amplificador operacional, $|A_{BA}|$, y de la realimentación, $1/\beta$, en la situación correspondiente al límite de estabilidad empleando la compensación de fase.

Por otra parte, la frecuencia de corte entre la realimentación y la curva de ganancia de un amplificador operacional viene dada por[3]:

$$f_2 = B\beta_{f_2} \tag{16.24}$$

donde β_{f2} es el valor de la realimentación a la frecuencia f_2 que, teniendo en cuenta la expresión (16.21), es aproximadamente igual a $C_r/(C + C_r)$. Por tanto:

$$f_2 = B\frac{C_r}{C + C_r} \tag{16.25}$$

Igualando las ecuaciones (16.23) y (16.25), se obtiene que el condensador C_r, que corresponde al límite de estabilidad es:

$$C_r = \frac{1}{4\pi R_r B}\left(1 + \sqrt{1 + 8\pi R_r CB}\right) \tag{16.26}$$

[3] Si se dispone de un amplificador compensado internamente cuyo ancho de banda a ganancia unitaria es igual a B y cuya realimentación es igual a β, se puede demostrar fácilmente que la frecuencia a la que se corta la curva de ganancia en bucle abierto del operacional y la curva $1/\beta$ viene dada por el producto $B\beta_{fc}$ donde β_{fc} es el valor de la realimentación en la propia frecuencia de corte.

Cualquier condensador C_r de un valor mayor al calculado en la expresión (16.26) logrará la compensación de fase pretendida.

La compensación realizada presenta el inconveniente de limitar el ancho de banda del sistema en bucle cerrado. *Así, la tensión de salida* v_o, *en el circuito de la figura 16.16 viene dada por:*

$$v_o = \frac{-R_r i_f}{1 + R_r C_r s} \tag{16.27}$$

que presenta un polo a la propia frecuencia f_2 que determina el ancho de banda. Por tanto, el condensador de compensación C_r, limita el ancho de banda del sistema cuyo máximo valor viene dado por el mínimo condensador C_r que garantice la estabilidad, es decir, el calculado en (16.26).

EJERCICIO 16.2

Diseñar un amplificador corriente-tensión de $10 \cdot 10^6$ *V/A y un ancho de banda de 1 kHz, sabiendo que se pretende utilizar un amplificador operacional de 1 MHz de ancho de banda a ganancia unitaria y que el fotodiodo tiene una capacidad de 100 pF.*

SOLUCIÓN:

Puesto que el factor de conversión tensión-corriente deseado es de $10 \cdot 10^6$ V/A, la resistencia de realimentación, R_r, necesaria, será de 10 MΩ.

El condensador, C_r, para realizar la compensación de fase determina el ancho de banda del circuito. Su valor será teniendo en cuenta la expresión 16.23:

$$C_r = \frac{1}{2 \cdot \pi \cdot 1 \cdot 10^3 \text{Hz} \cdot 10 \cdot 10^6 \Omega} = 16 \text{ pF}$$

Falta comprobar si el circuito queda compensado con este valor de capacidad. Teniendo en cuenta la expresión (16.26), el condensador límite es de 1,3 pF. Luego con el condensador elegido se realiza correctamente la compensación de fase.

16.1.5.3 Problemática del amplificador corriente-tensión

Hasta el momento se han comentado algunos de los límites del amplificador corriente-tensión para fotodiodos como los problemas derivados de las corrientes de polarización, la necesidad de la compensación y la limitación en el ancho de banda que esto supone. Sin embargo, quedan aún por mencionar algunas cuestiones de índole práctica relacionadas con la resistencia de realimentación y las capacidades parásitas.

Resistencia de realimentación

Como ya se ha mencionado, la resistencia de realimentación puede ser, en algunos casos, de un valor muy elevado. En la práctica, puede resultar difícil encontrar en el mercado resistencias de valor mayor que algunas decenas de megaohmios. Para resolver el problema se puede recurrir a la utilización de "redes en T" como se muestra en la figura 16.18.

En este caso, la tensión de salida, v_o, es igual:

$$v_o = -\left[R_1 \left(1 + \frac{R_2}{R_3} \right) + R_2 \right] i_f \tag{16.28}$$

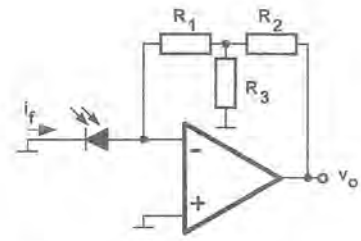


Figura 16.18. Esquema del amplificador corriente-tensión empleando una "red en T" en la realimentación.

Es decir, el conjunto de la red se comporta como una resistencia equivalente, R_{eqv}, de valor

$$R_{eqv} = R_1\left(1+\frac{R_2}{R_3}\right)+R_2 \qquad (16.29)$$

Con resistencias relativamente pequeñas se puede obtener una resistencia equivalente elevada. Por ejemplo, si $R_1 = R_2 = 1\ M\Omega$ y $R_3 = 10\ k\Omega$, la resistencia equivalente es igual a 102 MΩ. Sin embargo, la utilización de las "redes en T" presenta importantes inconvenientes, como son la amplificación de la tensión de desviación del operacional y el aumento del ruido interno [16.6], [16.10]. En la referencia [16.6] se puede ver el análisis detallado de estas cuestiones.

Capacidades parásitas

El valor necesario de la capacidad C_r, para realizar la compensación de fase puede ser muy pequeño cuando se requiera una ganancia y un ancho de banda elevados. En estos casos, las capacidades parásitas debidas a la resistencia de realimentación y los terminales de entrada y salida del operacional C_{Rpar} (figura 16.19a) pueden ser comparables o incluso mayores que el valor necesario para C_r. Estas capacidades son difícilmente predecibles aunque, en la práctica, la suma suele variar entre 0,1 y 0,5 pF. En ocasiones, estos valores son suficientes para lograr la compensación del circuito aun sin añadir la capacidad externa C_r. En otros casos, resultan excesivos y se ha limitado el ancho de banda innecesariamente.

Una solución muy sencilla para disminuir su efecto consiste en situar en la realimentación un conjunto de resistencias en serie cuyo equivalente sea igual a la resistencia deseada (figura 16.19b); pero, en algunos casos, esta solución puede ser insuficiente. En las referencias [16.2] y [16.6] se muestran otras soluciones basadas en la utilización de redes en T para la resistencia de realimentación.

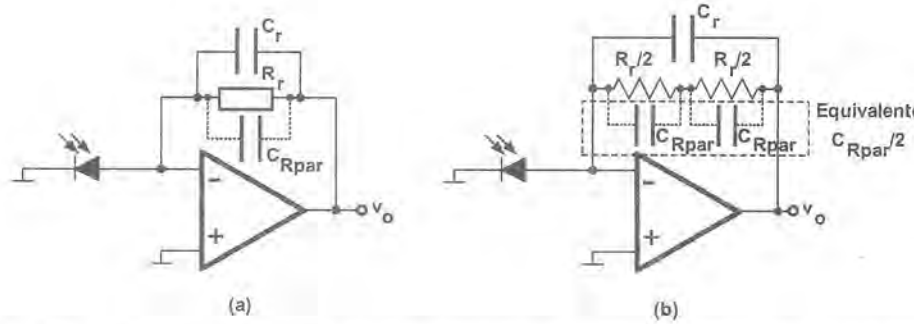


Figura 16.19. (a) Capacidad parásita en la realimentación debida la resistencia y a los terminales del operacional; (b) disminución de la capacidad parásita efectiva mediante asociación de resistencias en serie.

Por otra parte, las capacidades de modo diferencial C_{md}, y de modo común C_{mc}, del propio operacional se suman a la de transición del fotodiodo puesto que están en paralelo (figura 16.20). Estas capacidades son del orden de unos pocos picofaradios en la mayor parte de los operacionales y es necesario tenerlas en cuenta cuando la capacidad de transición del fotodiodo sea comparable o menor.

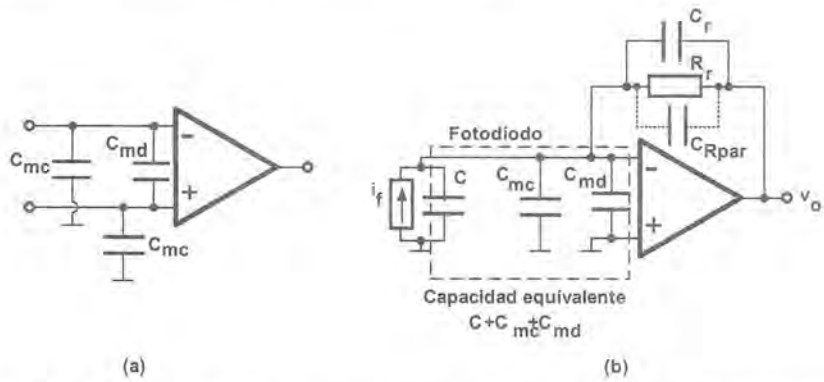


Figura 16.20. (a) Modelo del operacional incluyendo las capacidades en modo diferencial C_{md}, y en modo común C_{mc}; (b) circuito equivalente del convertidor corriente-tensión incluyendo las capacidades parásitas del operacional.

16.1.6 FOTOTRANSISTORES

Tanto la estructura interna de un fototransistor como su funcionamiento son similares a las de un transistor bipolar salvo que la inyección de la corriente de base se va a lograr, en el caso del fototransistor, mediante el mecanismo fotoeléctrico. Cuando incide un fotón con suficiente energía en la zona de transición constituida entre el colector y la base, se genera un par electrón-hueco. El campo eléctrico existente en la zona separa el par y acelera el electrón hacia la zona N (el colector) y el hueco hacia la zona P (la base) constituyendo una fotocorriente primaria, I_{fp} (figura 16.21a). El mecanismo de generación de esta fotocorriente es, por tanto, idéntico al de los fotodiodos. A partir de aquí, el principio de funcionamiento es idéntico al de un transistor bipolar.

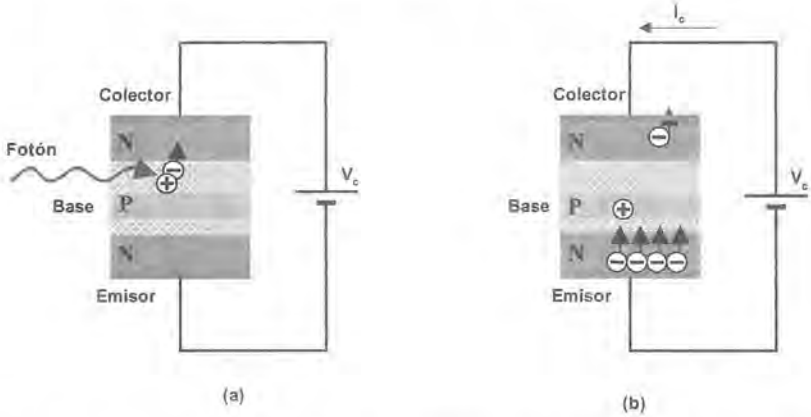


Figura 16.21. (a) Generación de un par electrón-hueco al incidir un fotón con energía mayor que la banda prohibida; (b) el hueco, una vez que ha alcanzado la zona N provoca la inyección de un elevado número de electrones para cancelarlo. Todos, menos uno, alcanzan el colector.

En definitiva, la corriente que circula por el colector es mucho mayor que la fotocorriente primaria y la relación entre ambas es la ganancia, β_T.

$$i_c = \beta_T i_{fp} \tag{16.30}$$

Las curvas características de un fototransistor (corriente de colector-tensión colector emisor) son similares a las de los transistores bipolares salvo que, en vez de reflejarse en función de la corriente de base, aparecen en función de la irradiancia de la luz incidente (figura 16.22).

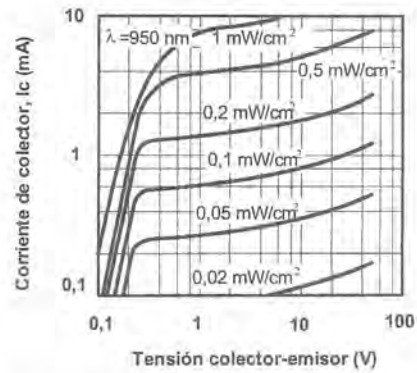


Figura 16.22. Curvas características del fototransistor BPW77N de Vishay para una longitud de onda de luz incidente igual a 950 nm.

La ventaja del fototransistor frente al fotodiodo es la ganancia. Sin embargo, presenta notables inconvenientes como la mayor dependencia de la temperatura, la menor linealidad y, sobre todo, la escasa reproducibilidad y el mayor tiempo de respuesta. Por ello, el fototransistor sólo se suele emplear para la detección de nivel de luz como sucede, por ejemplo, en las aplicaciones digitales. Para estas aplicaciones, los circuitos de acondicionamiento habituales son muy simples (figura 16.23), trabajándose, generalmente, entre la situación de corte (al incidir luz) y saturación (cuando no incide) o bien en zona activa cuando se necesita un tiempo de respuesta más reducido y añadiendo un comparador para discriminar el nivel.

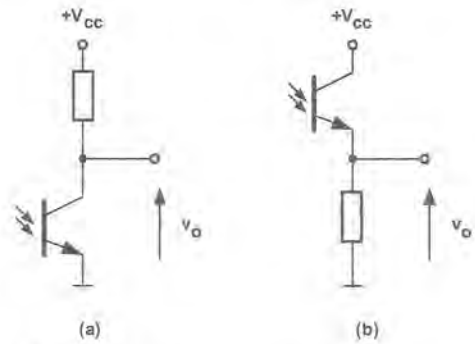


Figura 16.23. Circuitos de acondicionamiento para fototransistores.

16.1.7 APLICACIONES DE FOTODIODOS Y FOTOTRANSISTORES

En términos generales, las aplicaciones ópticas se pueden clasificar en dos tipos:

1. Aquellas en que la propia fuente de la luz a medir contiene la información. Es el caso, por ejemplo, de los fotómetros.
2. Aquéllas basadas en la medida de la modificación de la luz procedente de la fuente (generalmente, un LED) en función de la magnitud a medir. Esta alteración puede deberse a diferentes principios. Entre los más comunes se encuentran la absorción de la luz, la reflexión, la dispersión, la refracción y la interferencia con otro haz de luz.

En cualquiera de los casos, la luz proveniente del ambiente es una fuente de interferencias (salvo que sea el objeto de la medida). Para reducir su influencia se puede recurrir a las siguientes soluciones:

a) Proteger el sistema óptico de la luz ambiente cuando sea posible o, al menos, filtrar la luz que llega al detector mediante un filtro óptico que deje pasar únicamente las longitudes de onda de interés.

b) Modular la fuente de luz (figura 16.24). En este caso, se puede discriminar la luz proveniente de la fuente de la luz del ambiente mediante un filtrado electrónico. De todas formas, también suele ser necesario un filtrado óptico puesto que la luz interferente puede ser suficientemente elevada como para saturar el sistema de detección. La modulación de la fuente de luz es una solución que, normalmente, sólo se puede utilizar en el segundo tipo de aplicaciones anteriormente mencionado.

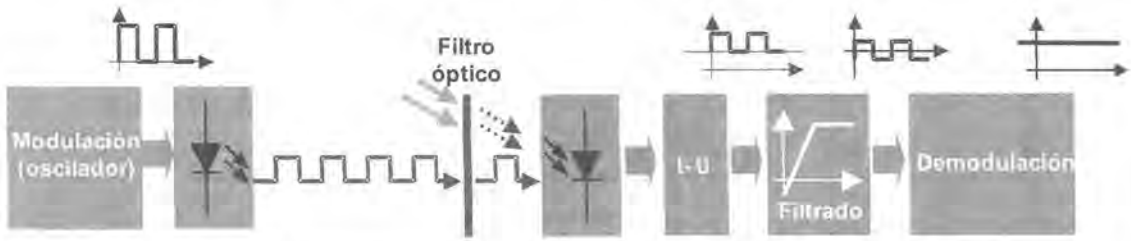


Figura 16.24. Diagrama de bloques típico de un sistema de medida de óptico.

Por sus excelentes características en cuanto a linealidad, ruido y velocidad de respuesta, el fotodiodo es el sensor de luz más utilizado en un buen número de aplicaciones ópticas, tanto analógicas como digitales. El fototransistor únicamente suele utilizarse para detectar niveles de luz (aplicaciones digitales) cuando no se requiere un tiempo de subida pequeño. Algunas de las aplicaciones típicas son las siguientes:

- Detectores de proximidad.
- Detectores de color.
- Detectores de humo.
- Turbidímetros.
- Instrumentación analítica basada en fluorescencia, absorbancia, colorimetría, bioluminiscencia.
- Detectores de lluvia en automóviles.
- Sistemas de comunicaciones por fibra óptica.
- Optoacopladores.

A continuación, se describen algunas de las aplicaciones más representativas.

16.1.7.1 Detectores de proximidad fotoeléctricos

Los detectores de proximidad ópticos están compuestos de una fuente de luz (un LED, generalmente) y un fotodetector. Existen, comercialmente, tres tipos fundamentales en función del modo de detección (figura 16.25):

- Detectores por reflexión en objeto (difusos): el LED y el fotodetector están montados sobre la misma cabeza de forma que en presencia de un objeto parte de la luz procedente del LED se refleja en el mismo y alcanza al fotodiodo. El alcance máximo de estos detectores es de unos 2 metros, aunque lo más habitual suelen ser algunos centímetros.
- Detectores por reflexión en espejo: un espejo refleja la luz del LED hacia el fotodetector, que también está situado en la misma cabeza, constituyéndose una barrera luminosa. La detección de un objeto se realiza a partir de la interrupción de la barrera. El alcance máximo es de unos 10 metros.
- Detectores de barrera: para alcances mayores se sitúa la fuente de luz y el fotodetector en extremos opuestos. El alcance puede llegar hasta los 50 metros con LEDs y hasta los 200 metros utilizando diodos láser como fuente luminosa.

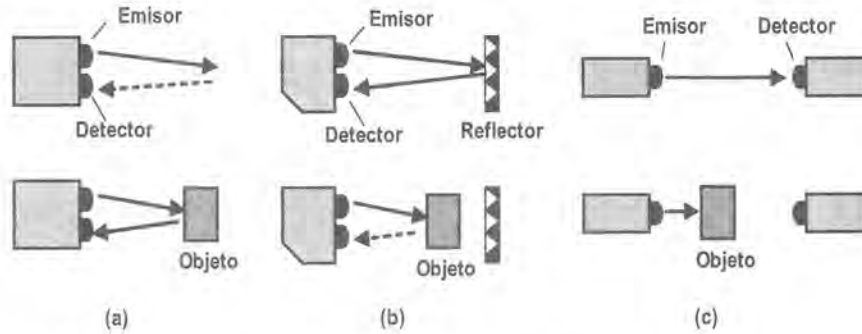


Figura 16.25. Principio de funcionamiento de diferentes tipos de detectores de proximidad fotoeléctricos: (a) detector de proximidad por reflexión en objeto; (b) detector de proximidad por reflexión en espejo; (c) detector de proximidad de barrera.

16.1.7.2 Codificadores ópticos (*optical encoders*)

Los codificadores ópticos se utilizan para la medida de posición y velocidad angular y lineal. Se clasifican en dos tipos, incrementales y absolutos, que se describen a continuación.

Codificadores ópticos incrementales

Están compuestos por un disco o un elemento lineal que se desplaza de forma solidaria con el elemento cuya posición se desea determinar y que está dividido en sectores con características ópticas diferentes, por ejemplo, transparentes y no transparentes, blancos y negros, u otras [16.11] [16.12]. Los más comunes están compuestos de un disco opaco ranurado; el paso de un hueco se detecta mediante un LED y un fotodetector enfrentados y situados a ambos lados del disco (figura 16.26). Estos dos elementos se suelen suministrar comercialmente en un único encapsulado bajo el nombre de fotointerruptores.

La medida de la velocidad se puede efectuar a partir de la medida de la frecuencia de la señal recibida en el fotodetector. Para realizar la medida de la posición basta contar el número de pulsos recibidos a partir de una referencia que suele venir determinada por una ranura adicional (ranura índice) cuyo paso se detecta mediante otro fotodetector (figura 16.27).

La resolución viene determinada por el número de ranuras o líneas del disco. Los de plástico y los metálicos están compuestos de hasta dos líneas por milímetro y 6 líneas/mm, respectivamente [16.12]. Con discos de cromo-cristal se llega hasta a 100 líneas/mm con el inconveniente de que su coste es muy elevado y, además, son muy frágiles. Electrónicamente, se puede duplicar la resolución si en vez de realizar la medida a

partir del nivel de los pulsos se realiza a partir de la detección de los flancos puesto que por cada pulso existen dos flancos [16.13].

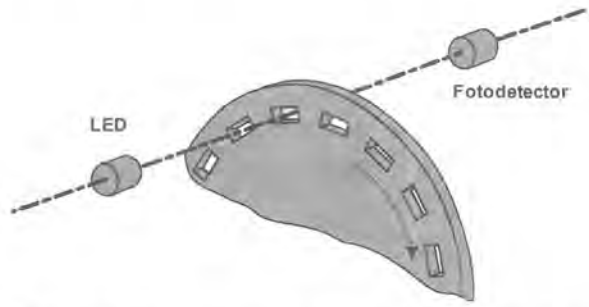


Figura 16.26. Codificador óptico compuesto por un disco opaco ranurado, un LED y un fotodetector situados a ambos lados del disco para la detección del paso de una ranura.

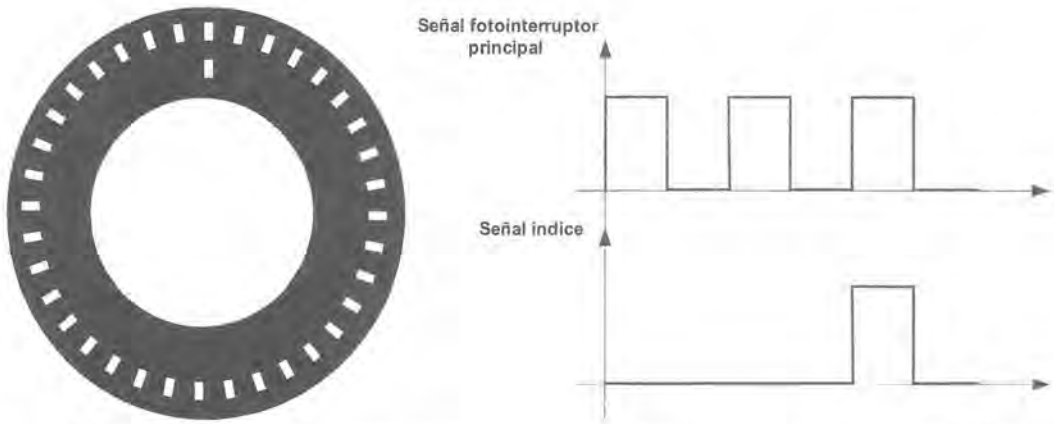


Figura 16.27. Medida de posición con un codificador incremental mediante la cuenta del número de pulsos a partir de una señal índice de referencia.

En muchas aplicaciones resulta necesario determinar el sentido de giro. La solución más habitual y económica consiste en situar dos fotodetectores de forma consecutiva (figura 16.28), de modo que cuando uno está situado enfrente de un hueco el otro está tapado. Las señales recibidas por ambos son idénticas pero una de ellas está adelantada respecto a la otra, dependiendo de cuál sea el sentido de giro. En la referencia [16.13] se puede consultar un método basado en la utilización de un biestable.

Codificadores ópticos absolutos

Mientras que en los codificadores incrementales la medida de la posición se realiza mediante de la cuenta de pulsos o flancos a partir de una señal de referencia, en los codificadores absolutos la medida de la posición se realiza directamente. Para ello, el disco está dividido en sectores codificados digitalmente según un código binario, Gray, BCD u otro (figura 16.29) que informa, directamente, de la posición.

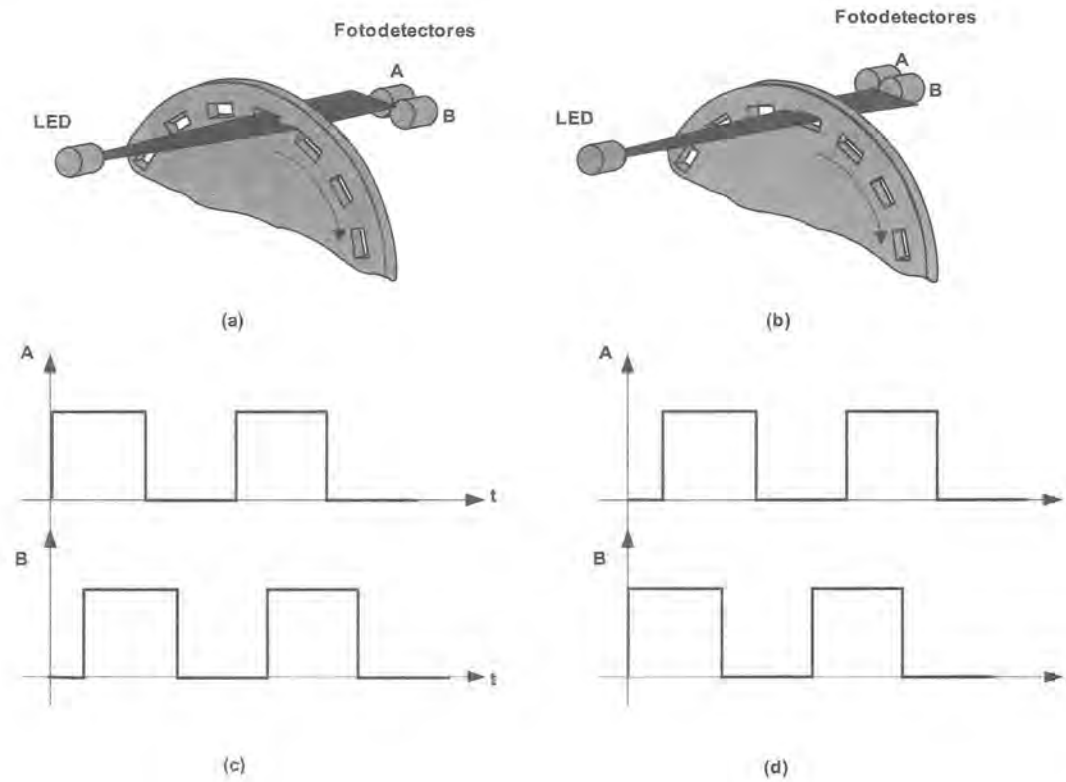


Figura 16.28. Detección del sentido de giro con un codificador incremental. (a) El fotodetector A recibe la luz procedente del LED mientras que el B está tapado; (b) al avanzar el disco en sentido horario, el hueco se aproxima a B, que recibe luz del LED, mientras que el A está tapado; (c) señales recibidas en los fotodetectores al girar el disco en sentido horario; (d) señales recibidas si el sentido es antihorario.

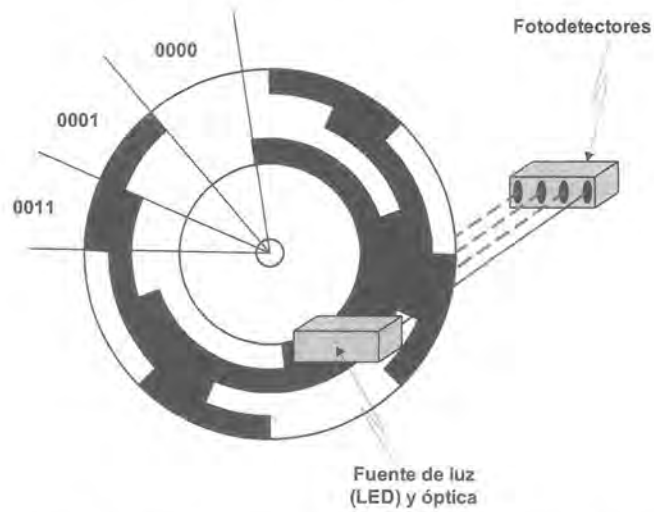


Figura 16.29. Principio de funcionamiento de un codificador óptico absoluto. En el ejemplo de la figura el disco está codificado en código Gray.

16.1.7.3 Sensores de color

Los sensores de color son muy utilizados, industrialmente, para la inspección y realización de controles de calidad (por ejemplo, en la industria alimenticia o para la detección de piezas previamente marcadas con etiquetas de diferentes colores en función de sus características). Los modelos más habituales contienen tres LEDs (uno azul, uno verde y uno rojo) que se van encendiendo de forma secuencial y cuya luz se dirige al objetivo (figura 16.30). El fotodiodo recibirá parte de la luz reflejada y tanta más cuanto más próximo esté el color del objetivo al de la luz emitida: por ejemplo, si el objetivo es de color azul quiere decir que absorbe todas las longitudes de onda de la zona visible salvo las correspondientes a la región azul del espectro, que las refleja; por tanto, el fotodiodo recibirá mucha luz azul, muy poca verde y apenas luz roja con lo que se puede identificar el color mediante una gradación de los colores.

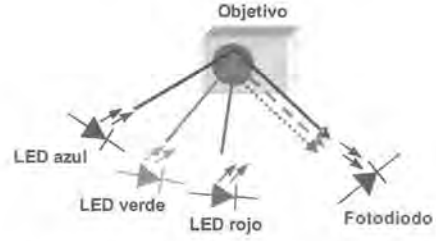


Figura 16.30. Principio de funcionamiento de un detector de color. Los LEDs se van encendiendo de modo secuencial: el fotodiodo recibe la luz correspondiente reflejada en el objetivo y que dependerá de su color.

16.1.7.4 Detectores de humo y turbidímetros

Aunque existen diversas alternativas para realizar la detección de humo, una de las más habituales consiste en realizar una detección de tipo óptica. En este caso, los detectores están compuestos por un LED y un fotodiodo situado formando un ángulo de 90° (figura 16.31) de forma que en ausencia de humo la luz del LED no incide en el fotodetector (figura 16.31a); por el contrario, en presencia de partículas de humo parte de la luz es dispersada y alcanza al fotodiodo (figura 16.31b).

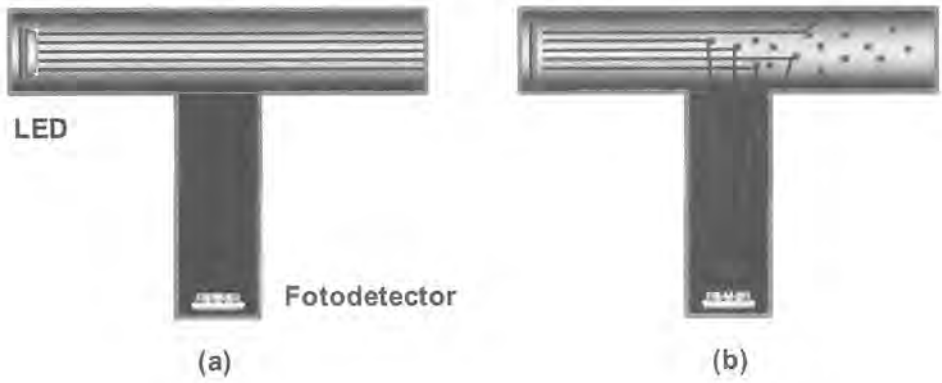


Figura 16.31. Principio de funcionamiento de un detector de humo.

Los turbidímetros son instrumentos para la medida de la turbidez del agua o cantidad de partículas sólidas disueltas. Su principio de funcionamiento es idéntico al de los detectores de humo. Generalmente, la medida se suele realizar por relación a la luz transmitida para lo que se sitúa otro fotodiodo enfrentado con la

fuente de luz; de este modo se reducen los efectos debidos a la variación de la fuente de luz y de la absorción del líquido. La aplicación fundamental de los turbidímetros es la realización de controles de calidad del agua. En los últimos años se ha convertido en un instrumento muy popular debido a su incorporación a algunos lavavajillas domésticos donde se utiliza para determinar el grado de suciedad del agua que vierten a los desagües y, de este modo, adaptar la duración del lavado y la cantidad de agua a utilizar para los aclarados a las necesidades concretas de cada ciclo en ejecución.

16.1.7.5 Espectrofotometría de absorción

La espectrofotometría de absorción es una técnica analítica que mide la cantidad de radiación electromagnética absorbida por un compuesto a una longitud de onda determinada. Aun cuando la radiación electromagnética se extiende en un margen espectral muy amplio, la porción de este espectro que va a considerarse en el presente apartado es la zona comprendida entre la radiación ultravioleta y el infrarrojo cercano.

Los espectrofotómetros se componen de una fuente luz capaz de emitir en las longitudes de onda objeto de estudio, un monocromador o elemento que permite seleccionar una estrecha banda espectral por medio de un prisma o una red de difracción y una rejilla de salida, un compartimento para colocar la muestra a estudiar y un fotodiodo (figura 16.32). Además, para evitar la incidencia de las variaciones de la luz incidente, la medida se suele efectuar por relación a esta. Para ello, se sitúa un divisor de luz entre el monocromador y la muestra y se mide la luz incidente mediante otro fotodiodo idéntico al anterior.

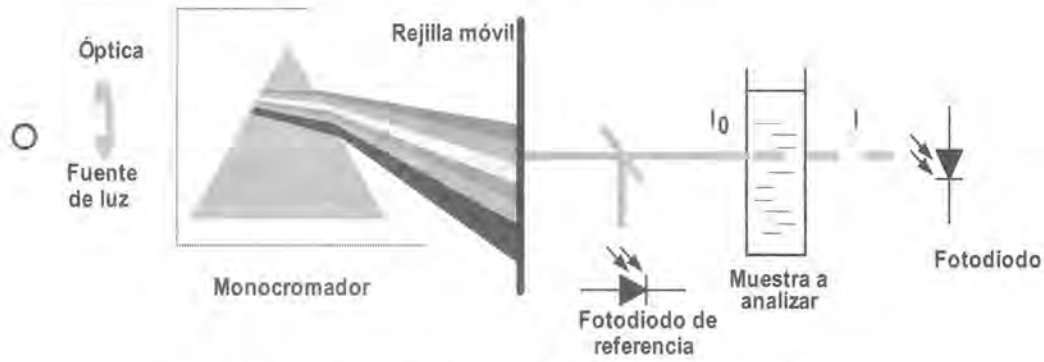


Figura 16.32. Esquema básico de un espectrofotómetro de absorción.

Si I_0 es la intensidad de radiación que incida en la muestra, la intensidad emergente, I, viene dada por (ley de Beer):

$$I = I_o \cdot 10^{-a \cdot b \cdot c} \tag{16.31}$$

donde a es una característica dada de cada sustancia bajo condiciones definidas de longitud de onda, temperatura, etc., b es la distancia recorrida por la radiación en el interior de la muestra y c es la concentración. Por otra parte, se define la absorbancia, A, como:

$$A = -\log \frac{I_o}{I} = a \cdot b \cdot c \tag{16.32}$$

Para una sustancia determinada, manteniendo constante el paso de la luz, las variaciones de concentración quedarán reflejadas en variaciones del valor de su absorbancia permitiendo realizar un análisis cuantitativo. Además, si la absorción depende de la naturaleza de la sustancia en estudio, se puede realizar un análisis cualitativo.

16.2 Dispositivos de acoplamiento de carga (*CCDs*)

16.2.1 PRINCIPIO DE FUNCIONAMIENTO

Los *CCDs* (*charge-coupled devices*) fueron desarrollados en el año 1970 por los laboratorios Bell. Inicialmente fueron concebidos como memorias analógicas. Sin embargo, sus aplicaciones fundamentales, hoy en día, son los sensores de imagen (por ejemplo, para cámaras fotográficas digitales, de vídeo, etc.) por lo que el estudio a lo largo del presente capítulo se realizará teniendo en cuenta esta aplicación y se utilizará el término CCD como sinónimo del sensor de imagen en conjunto.

En principio, un sensor de imagen podría estar constituido por una matriz de píxeles constituida por elementos fotosensibles (por ejemplo, fotodiodos) de forma que, empleando la óptica conveniente para enfocar el objeto sobre el sensor, se podría obtener la información de la imagen efectuando la lectura de cada uno de los píxeles (figura 16.33). Evidentemente, cuanto mayor sea el número de puntos mayor será la resolución de la imagen.

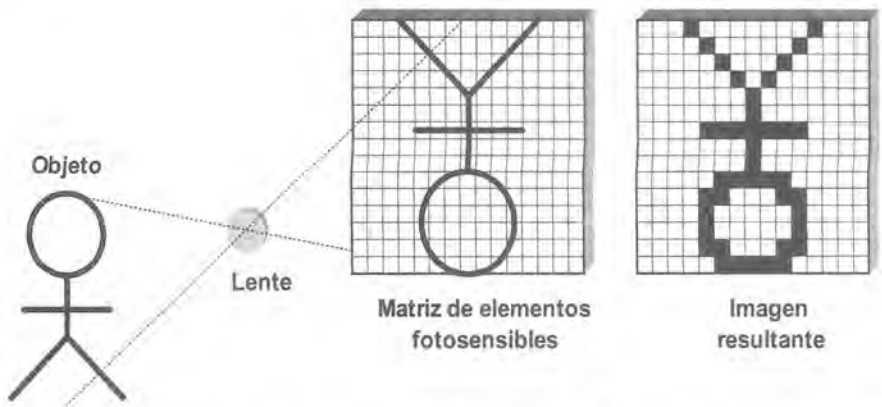


Figura 16.33. Obtención de una imagen de un objeto mediante una matriz de elementos fotosensibles.

Sin embargo, realizar la matriz de puntos simplemente con fotodiodos contaría con un importante inconveniente. Para obtener una imagen de buena calidad pueden necesitarse miles o incluso millones de píxeles. Un sensor de imagen constituido por fotodiodos y que tuviera accesible cada uno de los pines de los mismos sería, pues, inviable[4]. En los CCDs se resuelve este problema mediante un mecanismo que permite extraer la información de cada uno de los pixeles por un único pin en forma de una ristra de datos en serie.

En el funcionamiento de un CCD se distinguen tres fases diferentes (figura 16.34):

1. Conversión de la luz incidente a carga eléctrica.
2. Transferencia de las cargas.
3. Conversión de la carga a tensión (etapa de salida).

[4] Existen, comercialmente, matrices de fotodiodos (*photodiode arrays*) constituidos de hasta algunas decenas de fotodiodos. Su aplicación fundamental es la espectrometría.

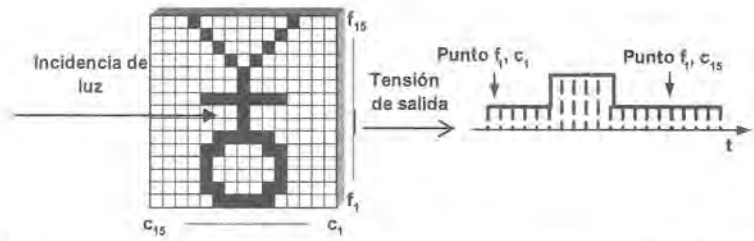


Figura 16.34. Esquema básico de funcionamiento de un CCD. La salida del CCD es una tensión cuyo valor a lo largo del tiempo se corresponde con la luz incidente en cada píxel.

16.2.1.1 Conversión de la luz a carga eléctrica

La estructura básica de cada píxel sensible de un CCD se muestra en la figura 16.35. Cada píxel está constituido por, al menos, tres electrodos separados de la zona activa por un aislante (óxido de silicio). Cuando un fotón incide en la zona del electrodo positivo, se puede crear un par electrón-hueco por el mismo mecanismo que el visto en el caso de los fotodiodos. El electrón queda atrapado en la zona del electrodo positivo puesto que lo atrae, mientras que los electrodos laterales tienden a repulsarlo y evitar que sea atraído por el electrodo positivo del píxel adyacente para que no se mezcle la información de los diferentes píxeles. Si sigue incidiendo la luz, se crean más electrones y se va produciendo una acumulación de carga.

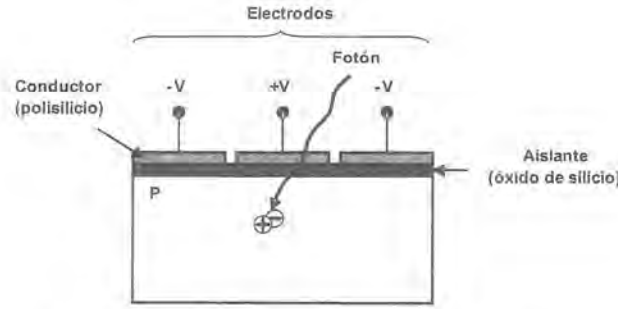


Figura 16.35. Esquema de un píxel de un CCD.

Cada píxel tiene, por lo tanto, las mismas limitaciones que un fotodiodo en cuanto a longitud de onda de corte, profundidad de penetración, etc. El modelo eléctrico se corresponde al de un fotodiodo en paralelo con un condensador (figura 16.36), que refleja la acumulación de la carga durante el tiempo que exponga el píxel a la luz hasta que se extraiga la información, denominado tiempo de exposición o tiempo de integración T_s.

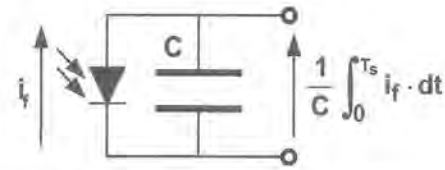


Figura 16.36. Modelo eléctrico de un píxel sensible. La capacidad mostrada trata de modelar el incremento de la carga en cada píxel con el tiempo de exposición a la luz.

16.2.1.2 Transferencia de las cargas

Una vez que el CCD ha sido expuesto durante un tiempo determinado a la luz, surge el problema de cómo mover la información de cada píxel a lo largo del dispositivo para llevarla al pin de salida. La técnica para realizar esta transferencia es el aspecto característico del CCD y de ahí su nombre. Existen las siguientes posibilidades, entre otras [16.4]:

- Método de las cuatro fases (4Φ).
- Método de las tres fases (3Φ).
- Método de las pseudo dos fases (P2Φ).
- Método de las verdaderas dos fases (T2Φ).

A continuación, se describen, brevemente cada una de ellas:

Método de las cuatro fases

Cada píxel está constituido por cuatro electrodos: dos corresponden a la zona fotosensible y los otros dos sirven para aislar cada píxel del adyacente y permitir la transferencia de la carga. Durante el tiempo de exposición de un píxel P_n, se mantienen a nivel alto los electrodos Φ1 y Φ2 y a nivel bajo los electrodos Φ3 y Φ4 (figura 16.37); a continuación, comienza la transferencia de carga de un píxel a otro cuyo procedimiento consta de cuatro pasos:

1. En primer lugar se invierten las polaridades de los electrodos Φ1 y Φ3. De este modo, el píxel avanza un electrodo hacia la derecha debido a la atracción electrostática de la carga que queda situada bajo los electrodos Φ2 y Φ3.
2. Seguidamente, se invierten las polaridades de Φ2 y Φ4, con lo que se vuelve a avanzar otro electrodo.
3. En tercer lugar, se invierten las polaridades de Φ3 y Φ1, quedando la carga bajo los electrodos Φ4 y Φ1.
4. Finalmente se invierten de nuevo las polaridades de Φ2 y Φ4 de modo que la carga quede situada bajo los electrodos Φ1 y Φ2 del píxel adyacente.

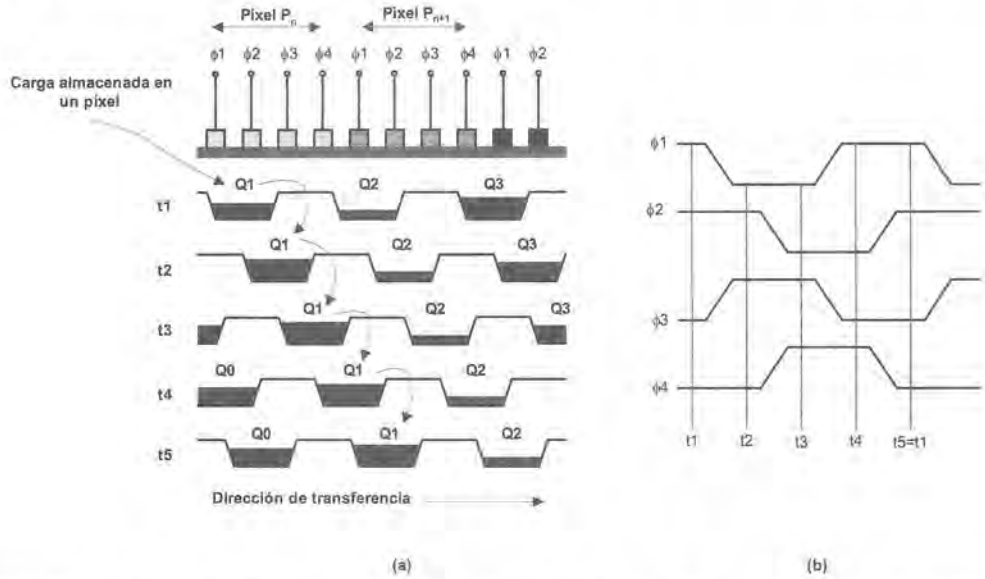


Figura 16.37. Método de las cuatro fases. (a) Estructura de cada píxel y evolución en el tiempo de la carga almacenada en cada píxel con las señales de gobierno de los electrodos mostradas en la figura (b) (tomado de [16.4]).

Como se puede observar en la figura 16.37, las señales de gobierno de los diferentes electrodos son idénticas, pero cada una de ellas va ligeramente retrasada sobre la anterior. Estas señales se obtienen a partir de osciladores (habitualmente, de cuarzo).

Método de las tres fases

En este caso, cada píxel está constituido por tres electrodos únicamente (figura 16.38). Durante el tiempo de exposición, la carga se almacena en el electrodo Φ1 manteniéndolo a nivel alto, mientras que Φ2 y Φ3 se mantienen a nivel bajo formando la barrera con el siguiente píxel. El proceso de transferencia se inicia poniendo a nivel alto el electrodo Φ2, con lo que la carga queda distribuida entre los electrodos Φ1 y Φ2. Poco tiempo después hay que poner Φ1 a nivel bajo, con lo que la carga queda únicamente bajo el electrodo Φ2. De un modo similar, se continúa actuando sobre los electrodos hasta que se complete la transferencia del píxel.

La ventaja del método de las tres fases sobre el de cuatro es que al ser menor el número de electrodos, la densidad de integración y la resolución son mayores. Presenta como inconveniente que debe prestarse especial cuidado en la obtención de las señales de reloj para obtener buenos resultados.

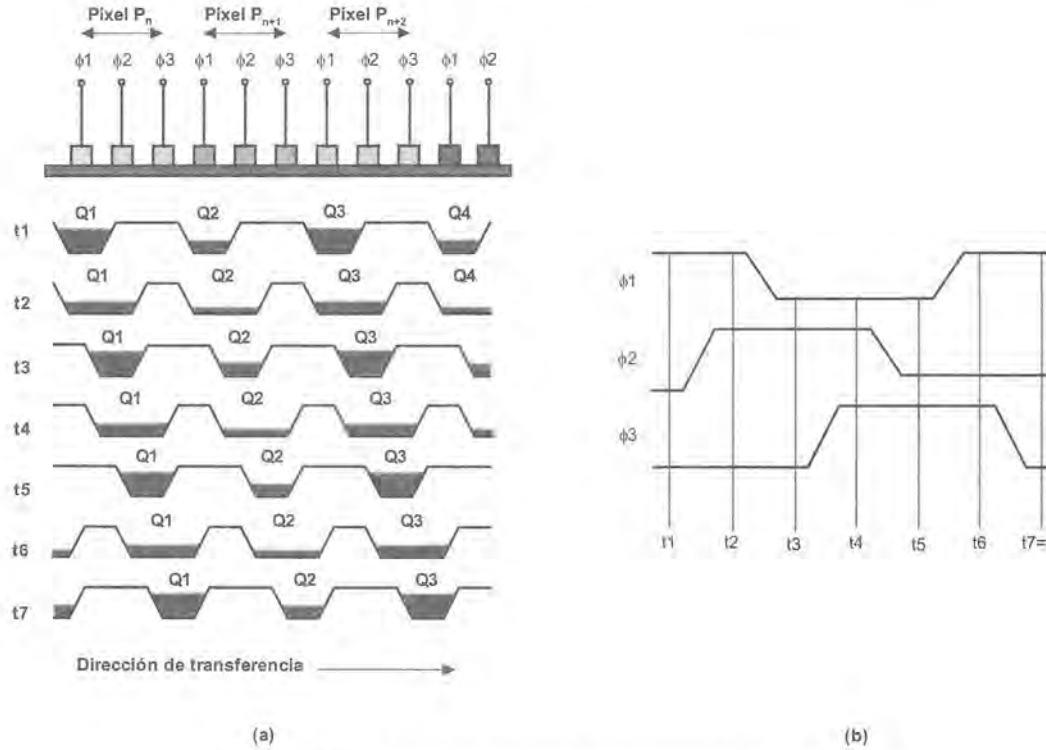


Figura 16.38. Método de las tres fases (tomado de [16.4]).

Método de las pseudo dos fases

Es un método muy similar al de las cuatro fases, aunque externamente tan sólo es necesario manejar dos señales de reloj que además están invertidas, lo que simplifica notablemente su obtención. Como se puede observar en la figura 16.39, cada señal de reloj gobierna dos electrodos. Para evitar que se mezcle la carga entre píxeles consecutivos, la estructura de cada uno de los electrodos que corresponden a una misma señal

de reloj es diferente, de forma que se logran campos eléctricos diferentes. Son dispositivos más costosos que los correspondientes a los métodos de tres o cuatro fases debido a que la estructura interna es más compleja.

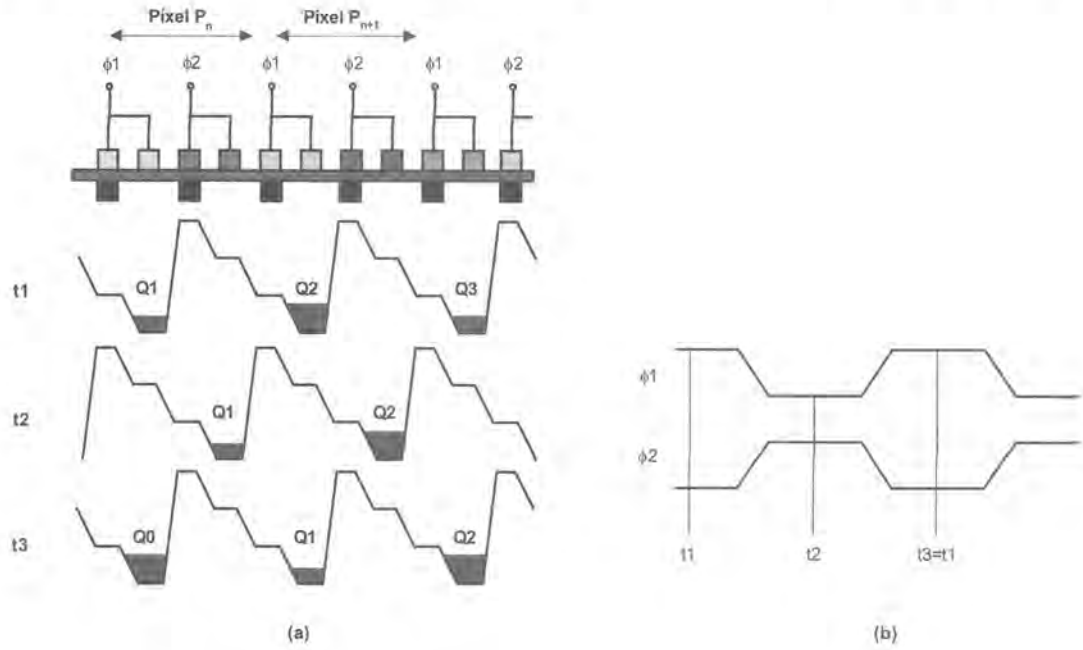


Figura 16.39. Método de las pseudo dos fases (tomado de [16.4]).

Método de las verdaderas dos fases

Se necesitan dos señales de reloj pero, a diferencia del método anterior, cada una de ellas sólo gobierna un electrodo. La estructura interna es tal que con únicamente dos señales se consigue un campo eléctrico escalonado en cada electrodo evitando la mezcla de la carga de los distintos píxeles. Además de su simplicidad de gobierno, presentan la ventaja de permitir densidades de integración mayores que el resto de los métodos.

16.2.1.3 Conversión de la carga a tensión (etapa de salida)

Mediante una adecuada transferencia se puede llevar la carga de cada píxel a un punto concreto de salida. Queda por resolver cómo transformar finalmente la carga en una tensión de salida proporcional que refleje el nivel de luz incidente en el punto.

En la figura 16.40 se muestra el esquema básico de la etapa de salida de un CCD [16.1]. La carga Q, de un píxel concreto se transfiere al condensador C_s, cuya tensión varía en el valor $\Delta V_{Cs} = Q/C_s$; la tensión de este condensador gobierna un transistor de tipo FET, configurado como seguidor, proporcionando la tensión proporcional a la carga sobre la resistencia de carga R_c. De todas formas, si la carga de todos los píxeles desembocara en el condensador C_s sin más precauciones, la tensión en el mismo iría acumulando todas las cargas, con lo que se mezclaría la información. Por ello, antes de transferir la carga al condensador es necesario fijar la tensión en el mismo a un valor de referencia cerrando el interruptor IN (*reset*).

Por tanto, el proceso es el siguiente (figura 16.41):

1. En primer lugar se cierra el interruptor de reset, IN. La tensión de salida toma el valor V_{o1}.
2. Seguidamente, se abre el interruptor IN. La tensión de salida toma el valor de referencia V_{ref}.

3. A continuación, se desplaza la carga al condensador. La tensión de salida toma el valor V_{o2}. La diferencia entre los niveles V_{ref} y V_{o2} es proporcional a la carga del píxel, que su a vez, es proporcional a la luz incidente en el mismo.

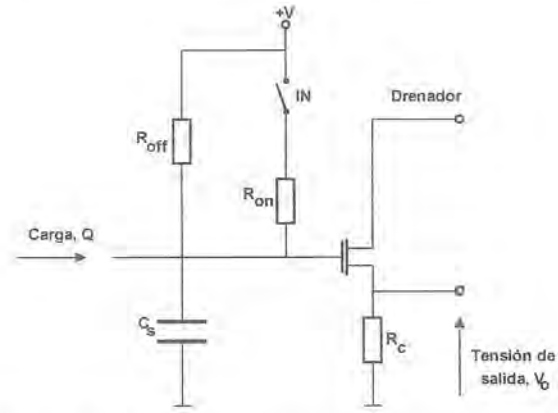


Figura 16.40. Esquema básico de la etapa de salida.

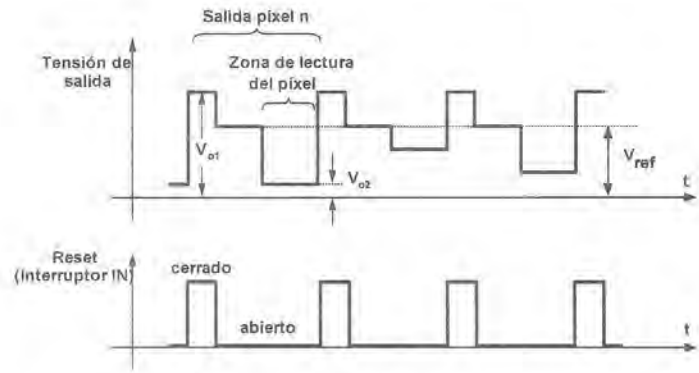


Figura 16.41. Tensión de salida de un CCD y actuación sobre el interruptor de reset.

Observe que la salida del CCD es analógica, es decir, la tensión $V_{ref} - V_{o2}$ es proporcional al nivel de luz incidente en el píxel. En muchas ocasiones, se supone que la tensión de referencia V_{ref}, es siempre la misma (el valor viene indicado en las hojas de características el dispositivo) con lo que únicamente se mide la tensión de salida después de haber efectuado la transferencia de la carga. Si se desea efectuar una lectura del pixel con mayor exactitud, será necesario medir tanto el valor exacto de V_{ref} para cada píxel como el valor de salida V_{o2} y realizar la resta. A esta técnica se le conoce como doble muestreo correlativo (*correlated double sampling, CDS*).

EJEMPLO 16.1

Los CCDs son dispositivos relativamente complejos, en lo que al gobierno se refiere, debido a la necesidad de múltiples señales para controlar la transferencia de las señales, el tiempo de exposición, etc.

Los fabricantes incluyen en las hojas de caracteristicas del dispositivo que se trate una serie de cronogramas cuyo perfecto entendimiento resulta crucial para cualquier realización práctica. Así, en la figura 16.42 se muestra el cronograma del CCD lineal SONY ILX533A. Se trata de un CCD lineal de 5150 píxeles efectivos. Como se puede observar en el cronograma, se necesitan 5 señales externas para gobernar el dispositivo:

- *La señal ϕROG controla el tiempo de exposición. Cuanto mayor sea el tiempo que se mantiene a este nivel mayor será el tiempo de exposición. Se pone a nivel alto cuando se desea realizar la transferencia de la imagen al registro de desplazamiento.*
- *Las señales $\phi 1$ y $\phi 2$ son las que gobiernan las transferencias de los pixeles a lo largo del registro de desplazamiento. La necesidad de dos únicas señales indica que se trata de un CCD que se controla por el método de las dos fases.*
- *La señal ϕRS corresponde al reset de la salida.*
- *Finalmente la señal ϕCLP gobierna la transferencia del punto de salida.*

Como se puede observar, en la tensión de salida están presentes no sólo los píxeles correspondientes a las zonas fotosensibles, sino que existen otros píxeles adicionales ciegos (dummy pixels) que corresponden al camino del CCD comprendido entre la salida y la zona de la transferencia.

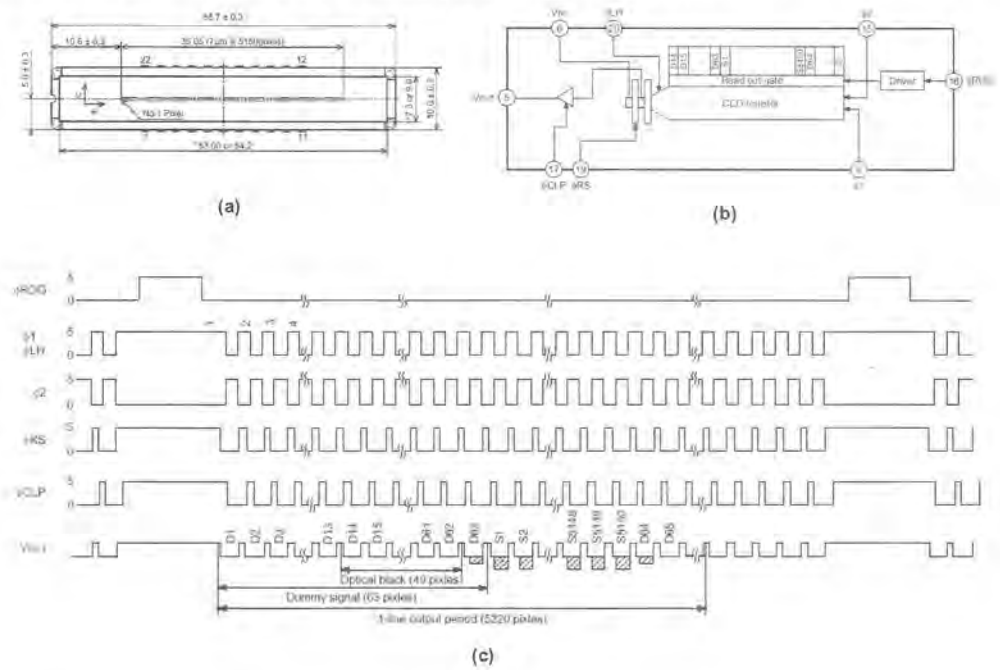


Figura 16.42. (a) Aspecto y dimensiones del CCD Sony ILX533A; (b) diagrama de bloques de la estructura interna; (c) cronograma (tomado de [16.15]).

16.2.2 ARQUITECTURAS

La arquitectura determina la disposición geométrica de los diferentes elementos sensibles y de los caminos para la extracción de la información. Se distinguen, fundamentalmente, dos tipos de CCDs: lineales y de superficie.

- Lineales (figura 16.43): están constituidos por una única línea de decenas o centenas de píxeles. Sus principales aplicaciones son los escáneres y los faxes. El dispositivo se expone a la luz durante el tiempo deseado; posteriormente se transvasa toda la información de los diferentes píxeles, mediante unos electrodos de transferencia, a un registro de desplazamiento (el verdadero CCD) protegido de la luz. De este modo, mientras se va sacando la información del registro de desplazamiento al exterior, se puede estar adquiriendo una nueva imagen en la zona fotosensible. Los CCDs bilineales incorporan dos registros de desplazamiento para aumentar la velocidad de transferencia.

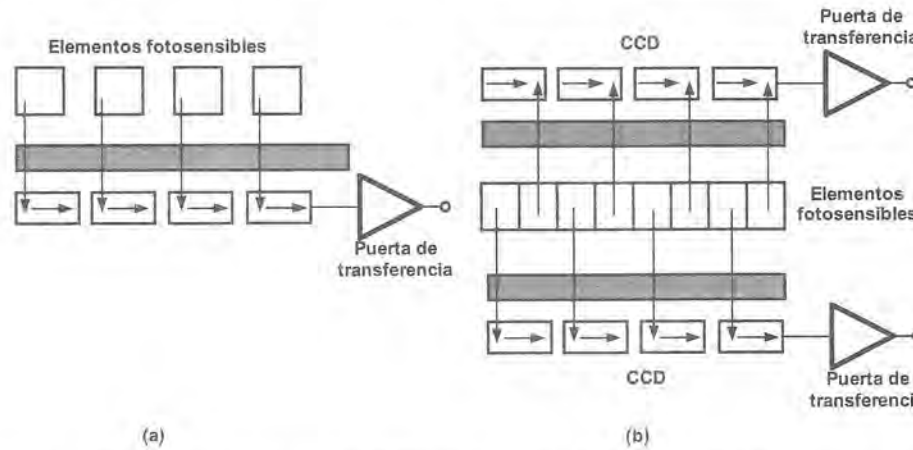


Figura 16.43. Arquitecturas de CCDs lineales: (a) CCD lineal; (b) de tipo bilineal.

- De superficie: se utilizan en cámaras de vídeo, fotográficas, etc. Dentro de ellos se distinguen tres tipos: CCDs de transferencia de línea, de interlínea o de transferencia de imagen [16.3], [16.5].
  - CCD de imagen de transferencia de línea (figura 16.44a): la matriz de elementos fotosensibles se expone a la luz ambiente durante el tiempo de integración necesario. A continuación, se desplaza toda la matriz una fila hacia abajo de forma que la última línea se transfiera a un CCD horizontal. Una vez que se ha sacado la información de toda la línea al exterior, se vuelve a desplazar la matriz una fila hacia abajo procediendo de igual modo y así sucesivamente hasta haber transferido toda la imagen. Puesto que mientras dura la transferencia la zona fotosensible queda expuesta, puede ser necesario taparla mediante un obturador mecánico hasta haber finalizado todo el proceso y estar en disposición de obtener una nueva imagen.
  - CCD de transferencia de imagen (figura 16.44b): el dispositivo está compuesto de dos matrices idénticas pero una está expuesta a la luz y la otra no. En primer lugar se realiza la transferencia completa de la primera a la segunda matriz, y posteriormente se procede a la transferencia de esta última al exterior mediante un procedimiento igual al CCD de transferencia de línea. Se utiliza mucho porque presenta la ventaja de que la transferencia entre matrices puede realizarse muy rápidamente evitando así la necesidad de obturador sin excesiva pérdida de calidad en la imagen y, además, mientras se está realizando la transferencia de la segunda matriz al exterior se puede estar adquiriendo una nueva imagen.
  - CCD de transferencia de interlínea (figura 16.44c): en este caso existen un registro de desplazamiento vertical por cada columna de elementos fotosensibles y un registro de desplazamiento horizontal. En primer lugar se realiza la transferencia completa de la carga almacenada en los elementos fotosensibles a los registros verticales, con lo que ya se puede proceder a la adquisición de una nueva imagen. Seguidamente, se procede al desplazamiento desde estos últimos al registro horizontal

línea a línea. Frente al CCD de transferencia de imagen tiene la ventaja de que el transvase a la zona no fotosensible se realiza más rápidamente puesto que sólo requiere un paso. El inconveniente es el peor aprovechamiento de la superficie del dispositivo y el consiguiente empeoramiento de la resolución.

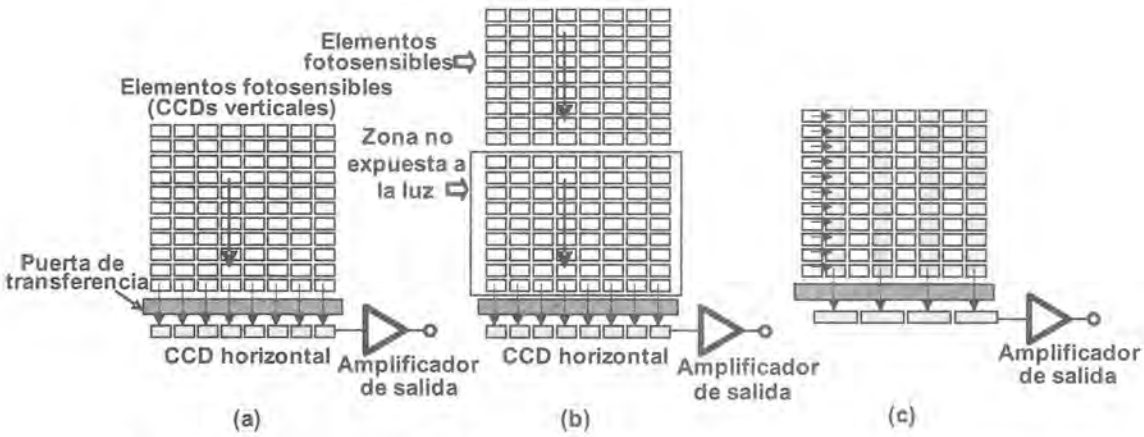


Figura 16.44. Arquitecturas de CCDs de superficie: (a) De transferencia de línea; (b) de transferencia de imagen; (c) de transferencia de interlínea.

16.2.3 CARACTERÍSTICAS

Las características fundamentales de los CCDs indicadas en los catálogos y sus significados son los siguientes [16.17]:

- Número de puntos (*number of pixels*): número de píxeles fotosensibles del dispositivo.
- Resolución (*resolution*): número de píxeles por unidad de superficie. Normalmente se expresa en puntos por pulgada (*dots per inch, DPI*).
- Respuesta espectral (*spectral response*): refleja, al igual que en los fotodiodos, la respuesta en función de la longitud de onda.
- Sensibilidad (*sensitivity*): puesto que la tensión de salida depende de la carga almacenada, y esta depende tanto de la luz que incide el píxel como del tiempo de exposición, los fabricantes suelen expresar la tensión de salida en función de ambos parámetros, habitualmente en V/lx·s.
- No uniformidad de la respuesta (*Photo Response non uniformity, PRNU*): es un parámetro para indicar la diferencia de la respuesta de los diferentes píxeles ante niveles idénticos de luz. Se define como:

$$\text{PRNU} = \frac{\Delta\chi}{\bar{\chi}} \tag{16.33}$$

  donde $\bar{\chi}$ es la media de las respuestas y $\Delta\chi$ es la máxima desviación respecto a la media.
- Nivel de tensión de referencia (*DC offset*): tensión de salida correspondiente al nivel de referencia.
- Tensión de oscuridad (*dark signal voltage*): variación de la tensión de salida correspondiente a un píxel con respecto al nivel de referencia cuando no incide luz sobre el mismo.

- Nivel de exposición de saturación (*saturation exposure*): cada píxel tiene una capacidad limitada para almacenar carga. Cuando se excede, el píxel queda desbordado y parte de la carga se transfiere a los píxeles vecinos "*blooming*" cuyas imágenes quedan distorsionadas. Los fabricantes indican el nivel de exposición que satura el dispositivo en lx·s, generalmente.
- Tensión de saturación (*saturation output voltage*): es la tensión de salida que corresponde a un píxel cuando está saturado.
- Eficiencia total de la transferencia de carga (*total charge transfer efficiency*): en la práctica, en el movimiento de un píxel se pierde una pequeña parte de la carga. En la transferencia completa a la salida la eficiencia del proceso suele rondar el 90-95%.

16.2.4 CCDs PARA DETECCIÓN DE IMÁGENES EN COLOR

Los CCDs destinados a aplicaciones en que resulta necesario obtener la información del color de la imagen (por ejemplo, en las cámaras de vídeo) presentan una estructura como la que se muestra en la figura 16.45a. Cada píxel efectivo está compuesto realmente de cuatro; cada uno de ellos incorpora un microfiltro óptico para fijar la respuesta espectral del píxel (figura 16.45b). Normalmente dos son de color verde, uno azul y el otro rojo. El funcionamiento, por otra parte, es idéntico al de los CCDs monocromos. La resolución efectiva en estos dispositivos se ve un tanto reducida. Por ello, en los sistemas de vídeo profesionales se emplean 3 CCDs, uno para cada color (figura 16.46), a costa de una complejidad y un coste notablemente mayores.

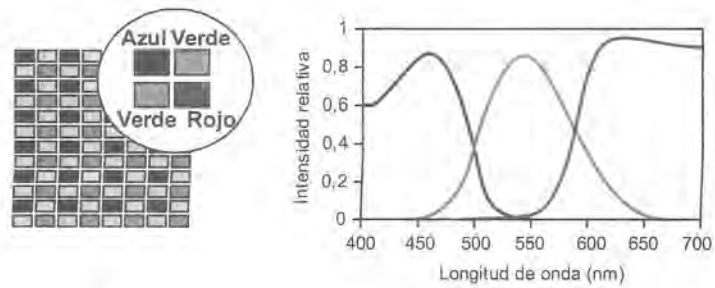


Figura 16.45. (a) Estructura de un CCD para color; (b) respuesta espectral de cada píxel.

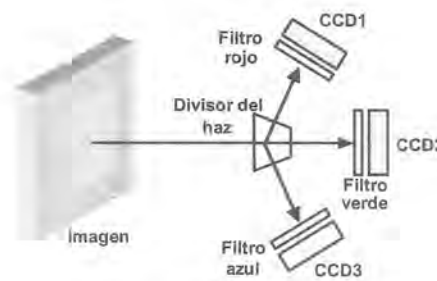


Figura 16.46. Detección de color mediante tres CCDs.

16.3 Fotomultiplicadores

16.3.1 PRINCIPIO DE FUNCIONAMIENTO

Un fotomultiplicador (*photomultiplier tube, PMT*) es un dispositivo que proporciona a la salida una corriente proporcional a la luz incidente. En la figura 16.47, se muestra la sección transversal. Los elementos fundamentales son el fotocátodo, los dínodos y el ánodo.

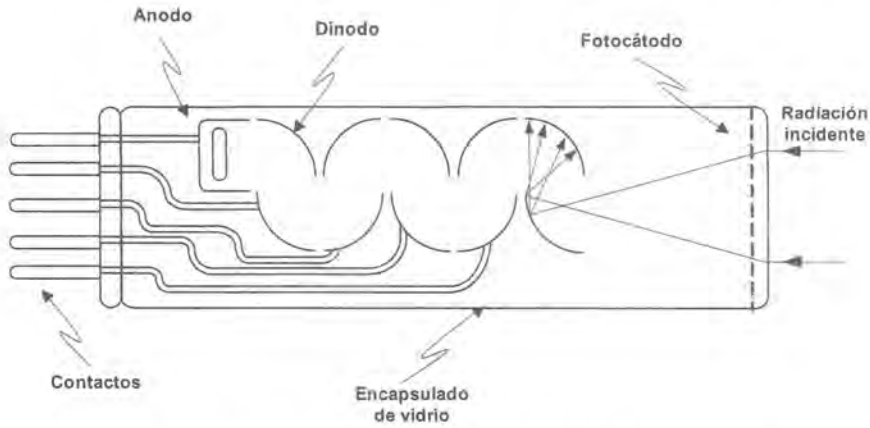


Figura 16.47. Sección transversal de un fotomultiplicador.

El funcionamiento de un tubo fotomultiplicador está basado en el efecto fotoeléctrico: cuando un fotón incide en el fotocátodo, que está recubierto con un material fotoemisor, este emite electrones que son acelerados y dirigidos a un multiplicador de electrones constituido por varios electrodos (denominados dínodos). Al incidir un electrón en uno de los dínodos, este emite varios electrones a su vez, que son acelerados hacia el siguiente dínodo gracias a que se aplica una diferencia de potencial entre ambos. Al final de proceso, el ánodo recoge todos los electrones obteniéndose a la salida una señal en forma de corriente, proporcional a la luz incidente.

Lo más característico del dispositivo es el efecto multiplicador o ganancia que se obtiene y que, además, puede ser tremendamente elevada.

16.3.2 TIPOS DE FOTOMULTIPLICADORES

Los fotomultiplicadores se clasifican en función de la situación de la ventana (el fotocátodo) y el multiplicador de electrones. Atendiendo a la ventana, cabe distinguir entre los de ventana lateral (generalmente, muy sensibles y rápidos) y los de ventana frontal (figura 16.48a).

Respecto al multiplicador de electrones existen, entre otros, los siguientes tipos [16.7] (figura 16.48b):

- Jaula circular (*circular cage*): utilizado principalmente en fotomultiplicadores de ventana lateral. Compactos y relativamente baratos, su tiempo de respuesta es pequeño.
- Caja y rejilla (*box and grid*): se utilizan en fotomultiplicadores de ventana frontal cuando el tiempo de respuesta no es un factor crítico.

- Enfocados linealmente (*linear focused*): presentan un tiempo de respuesta, ganancia y linealidad excelentes. Por el contrario, su tamaño es grande y su respuesta puede verse modificada por la presencia de campos magnéticos.
- Persiana veneciana (*venetian blind*): se utilizan para tubos con un fotocátodo de grandes dimensiones. Buenas características generales, excepto el tiempo de respuesta.

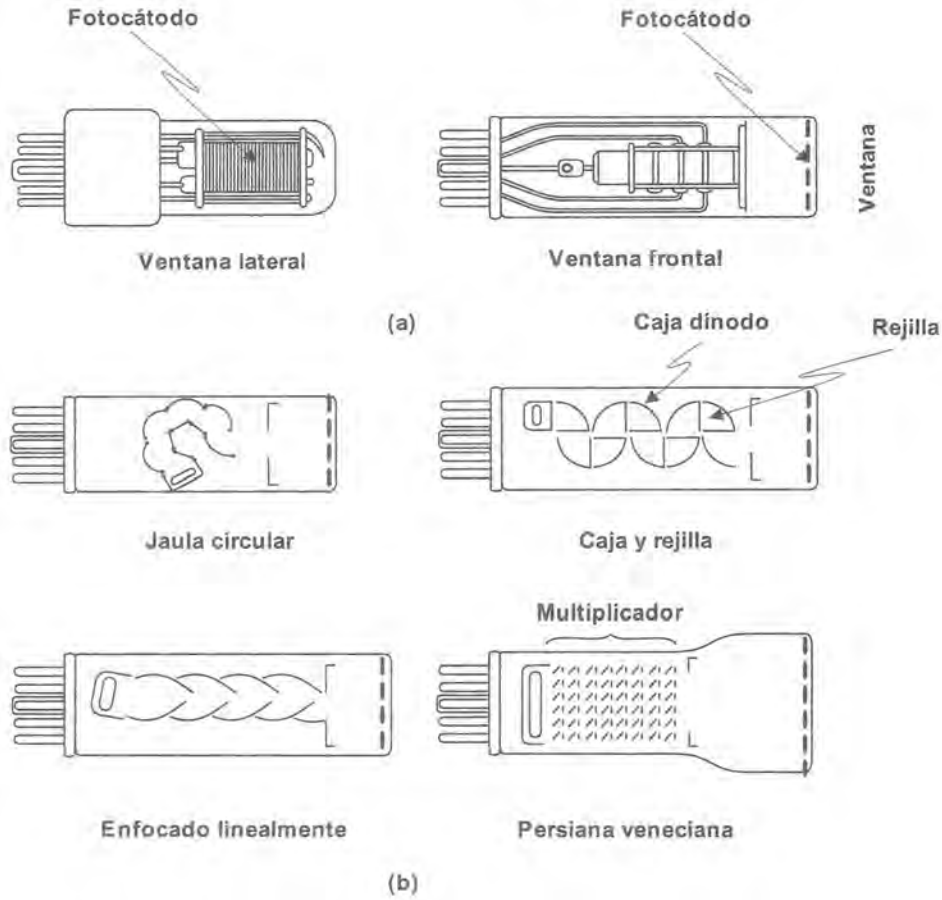


Figura 16.48. Tipos de fotomultiplicadores: (a) Atendiendo a la situación de la ventana; (b) atendiendo al multiplicador de electrones.

16.3.3 CARACTERÍSTICAS

Las características más importantes de selección de los fotomultiplicadores son las siguientes:

- Respuesta espectral (*spectral response*): es la sensibilidad del fotocátodo (corriente de salida dividida entre la potencia de la luz) en relación con la longitud de onda de la luz incidente. Observe que el término viene definido únicamente respecto al fotocátodo, es decir, sin tener en cuenta el efecto multiplicador. En ocasiones, los fabricantes indican en su lugar la eficiencia cuántica, cuyo significado es similar al ya indicado en el caso de los fotodiodos. El material del fotocátodo condiciona la respuesta espectral. Algunos materiales típicos son los siguientes [16.7]:

- Monoalcalinos (Sb-Cs): cubren desde el ultravioleta hasta el visible.
- Bialcalinos (Sb-Rb-Cs, Sb-K-Cs): presentan una respuesta espectral similar al anterior pero con mayor sensibilidad y menor ruido.
- Multialcalinos (Sb-Na-K-Cs): cubren desde la región ultravioleta hasta el infrarrojo. El mayor inconveniente es que la corriente de oscuridad es relativamente alta.

- Ganancia (*gain*): es la relación entre la corriente de ánodo y la corriente de fotocátodo. Depende de las características del fotomultiplicador y de la tensión de alimentación (figura 16.49) y puede ser tremendamente elevada (10^5-10^7 a la tensión nominal de alimentación).

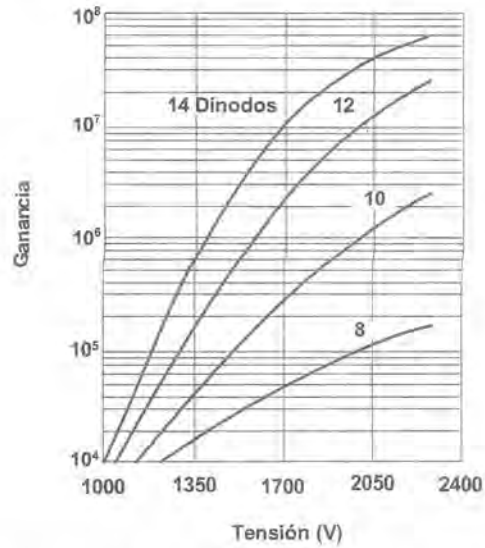


Figura 16.49. *Relación entre la tensión de alimentación y la ganancia, mostrando el efecto del número de dínodos (tomado de [16.16]).*

- Sensibilidad respecto al ánodo (*anode sensitivity*): es la corriente de ánodo por vatio de luz incidente. Es un parámetro que tiene en cuenta, por tanto, la respuesta espectral y la ganancia.
- Corriente de oscuridad (*anode dark current*): es la corriente de salida (en el ánodo) cuando no incide luz en el fotocátodo. Aumenta con la tensión de alimentación y con la temperatura. También depende del tipo de fotocátodo.
- Tiempo de respuesta (*time response*): es el tiempo que tarda la corriente de ánodo en pasar del 10% al *90% del valor final ante un escalón de luz incidente. Los valores típicos son de algunos nanosegundos.*
- Ruido equivalente a la entrada (*equivalent noise input, ENI*): es la potencia de luz incidente que provocaría a la salida del fotomultiplicador la misma corriente que la debida al ruido.

16.3.4 ALIMENTACIÓN

La tensión a aplicar entre los diferentes dínodos del fotomultiplicador se suelen obtener mediante un divisor resistivo a partir de una fuente de alimentación de alta tensión (entre *500* y *2000* V, en general).

El divisor de tensión viene condicionado por el tipo de aplicación. Cuando se desean medir niveles continuos de luz, *se suele aplicar una tensión igual en todas las etapas (figura 16.50a) y de forma que la corrien-*

te a través del divisor sea, al menos, 10 veces la corriente media de ánodo; de lo contrario tendría lugar una importante pérdida de linealidad. Si la aplicación consiste en medir niveles de luz pulsantes, el divisor se diseña de forma que la corriente por el mismo sea, como mínimo, 100 veces la corriente media de ánodo. Además, se suele aplicar una tensión más elevada en las últimas etapas para evitar la saturación del dispositivo debido a una elevada densidad de corriente fotoeléctrica y se añaden condensadores en paralelo con las resistencias de las últimas etapas para evitar variaciones de tensión (figura 16.50b).

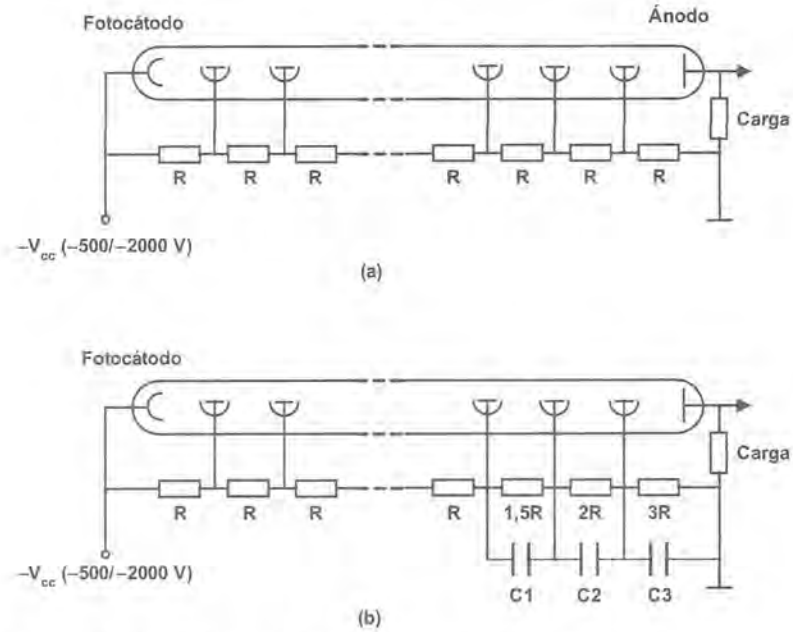


Figura 16.50. (a) Configuración del divisor resistivo para aplicaciones destinadas a la medida de luz continua; (b) posible configuración del divisor para la medida de fuentes de luz pulsantes.

Por otra parte, la ganancia del fotomultiplicador varía extraordinariamente con la tensión de alimentación, por lo que, en cualquier caso, deberá tener un rizado de tensión y derivas muy pequeños.

16.3.5 ACONDICIONAMIENTO DE SEÑAL

Puesto que un fotomultiplicador proporciona una corriente de salida proporcional a la luz incidente el modelo puede asimilarse a una fuente de corriente (como en un fotodiodo). El objetivo esencial del circuito de acondicionamiento será, por tanto, convertir la corriente a tensión. Una posibilidad es conectar simplemente una resistencia de carga R_c, a la salida del amplificador (figura 16.51).

El problema del circuito anterior es que la tensión entre el último dínodo y el ánodo varía con la caída de tensión en la carga, provocando una pérdida de linealidad en el dispositivo. Otro problema adicional es que, si la resistencia de carga tiene un valor óhmico elevado (téngase en cuenta que la corriente de salida suele estar comprendida entre 0,1 y 100 μA), la capacidad de salida del fotomultiplicador[5] sumada a la del propio cableado (que puede ser bastante mayor), puede limitar de forma muy importante la respuesta dinámica.

[5] Generalmente, los fabricantes no suelen indicar el valor de la capacidad de salida debido a que su valor es pequeño (<5 pF) y suele tener poca importancia en el diseño. Normalmente suele ser más limitativa la capacidad del propio cableado y la del amplificador.

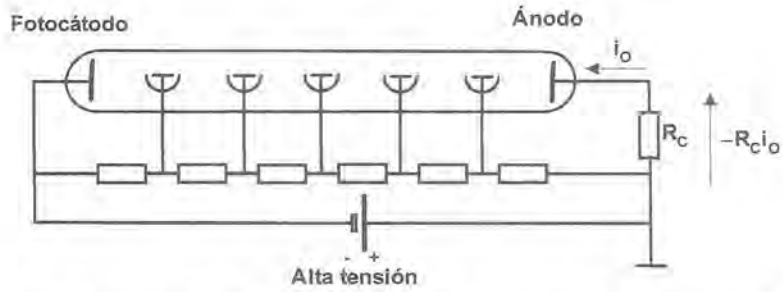


Figura 16.51. Conversión de la corriente a tensión mediante una resistencia de carga.

Otra posibilidad más habitual es emplear un amplificador corriente-tensión (figura 16.52), cuyo diseño se realiza de forma análoga al mostrado para los fotodiodos puesto que el modelo eléctrico del fotomultiplicador es muy similar.

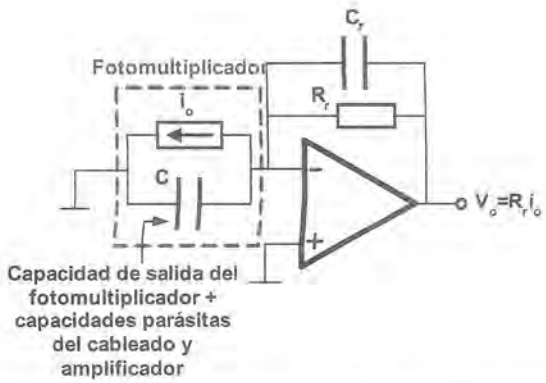


Figura 16.52. Circuito de acondicionamiento basado en la utilización de un amplificador corriente-tensión.

16.3.6 APLICACIONES

La ventaja de los fotomultiplicadores sobre los dispositivos de estado sólido (como fotodiodos) es la extraordinaria sensibilidad que poseen debido a la ganancia que ofrece el multiplicador de electrones. Los mayores inconvenientes del dispositivo son su elevado precio y tamaño, su fragilidad y la duración un tanto limitada (10^3-10^4 horas). El envejecimiento de estos dispositivos viene ocasionado, sobre todo, por el desgaste del último dínodo debido al bombardeo de electrones. Por ello, no conviene que la corriente media de salida sobrepase 1-2 μA.

Es un dispositivo hoy por hoy insustituible cuando el nivel de luz a medir es muy bajo. Algunas aplicaciones típicas son la espectrometría, los sistemas de medida remota de contaminación atmosférica mediante láser, astronomía y algunas aplicaciones médicas.

Resumen

Existen, fundamentalmente, cuatro tipos de fotodiodos: PN, PIN, Schottky y de avalancha. El fotodiodo PN es el más común dentro de los sistemas de instrumentación electrónica. El fotodiodo PIN y el de avalancha

se utilizan, sobre todo, en sistemas de comunicaciones. La principal virtud del fotodiodo Schottky es que presenta una respuesta mejor que los anteriores en la región ultravioleta.

El modelo eléctrico más común para un fotodiodo se compone al menos de una fuente de corriente proporcional a la luz incidente, un diodo en paralelo para tener en cuenta su comportamiento como tal diodo cuando se trabaja en zona directa y un condensador, también en paralelo, para tener en cuenta la capacidad de transición.

Algunas de las características más importantes para seleccionar un fotodiodo son la superficie y la respuesta espectral. Cuando se desean medir niveles de luz bajos también resultan interesantes la corriente de oscuridad y el ruido. El tiempo de subida y la capacidad son características a tener en cuenta cuando se desea una elevada rapidez.

Las mejores estrategias para medir luz con un fotodiodo pasan, en general, por medir la corriente inversa.

El circuito de acondicionamiento de señal más habitual para medir la corriente inversa en un fotodiodo es el amplificador corriente-tensión.

En el diseño del amplificador corriente-tensión es necesario tener en cuenta que el circuito puede ser inestable debido a la capacidad del fotodiodo. Para estabilizarlo se puede añadir un condensador en paralelo con la resistencia de realimentación.

Los fototransitores presentan, en general, peores características que los fotodiodos. Su aplicación fundamental es la detección de nivel y aplicaciones digitales poco exigentes.

En los CCDs se distinguen tres procesos fundamentales: la conversión de luz a carga eléctrica, la transferencia de cargas y la conversión de carga a tensión en la salida.

El CCD proporciona una tensión de salida proporcional a la luz incidente en cada píxel durante un tiempo determinado. Generalmente, la información de todos los píxeles sale en serie por un único pin de salida.

Se distinguen dos arquitecturas fundamentales de CCDs: de línea y de superficie. Dentro de estos últimos también se distinguen los CCDs de transferencia de línea, de transferencia de imagen y de transferencia de interlínea.

En los CCDs para la detección de imágenes en color, cada píxel efectivo está compuesto de varios píxeles con respuestas espectrales diferentes.

La característica más destacable de los fotomultiplicadores es la extraordinariamente elevada sensibilidad. Además son muy rápidos. Sus mayores inconvenientes son el precio, el envejecimiento y la fragilidad.

Los fotomultiplicadores proporcionan a su salida una corriente proporcional a la luz incidente por lo que los circuitos de acondicionamiento pueden ser similares a los de los fotodiodos.

Problemas propuestos

16.1 Se dispone de un fotodiodo de 1 mm^2 en el que incide una luz de 850 nm. Si la irradiancia de la luz incidente es de 1 $\mu W/cm^2$ y la corriente de cortocircuito de 5 nA, ¿cuál es la eficiencia cuántica a 850 nm?

16.2 Sobre un fotodiodo incide una onda cuadrada de luz procedente de un LED de 870 nm cuya frecuencia es de 1 kHz y sus niveles son 0 y 1 μW. El fotodiodo tiene una capacidad de 100 pF a 0 V

de tensión inversa y la sensibilidad es de 0,5 A/W a 870 nm. Diséñese un amplificador corriente-tensión que convierta el nivel de luz de entrada a una tensión, también cuadrada, entre 0-1 V. ¿Cuál debería ser el ancho de banda a ganancia unitaria del amplificador operacional?

16.3 Diséñese un circuito para convertir la tensión proporcionada por el convertidor corriente-tensión del problema anterior que proporcione a la salida una tensión continua comprendida entre 0-1 V, proporcional a la luz incidente en el fotodiodo.

16.4 Un CCD tiene un nivel de exposición de saturación de 0,14 lx·s. Si la iluminancia es de 100 lx, ¿cuál es el tiempo máximo de exposición permitido sin que se sature el dispositivo?

16.5 Si la sensibilidad de un CCD es de 15 V/lx·s, la tensión de referencia es de 5 V y la tensión de saturación es de 1 V, dibuje la tensión de salida correspondiente a un píxel cuando:

a) La iluminancia es de 10 lx y el tiempo de exposición de 0,005 s.
b) La iluminancia es de 1000 lx y el tiempo de exposición de 0,005 s.

16.6 Si la eficiencia cuántica del fotocátodo de un fotomultiplicador es del 50% para una longitud de onda de la luz incidente de 500 nm:

a) ¿Cuál es la sensibilidad a esa longitud de onda expresada en A/W?
b) Si además se alimenta a 1000 V y la ganancia del multiplicador es de 10^7 a esa tensión, ¿cuál es la correspondiente sensibilidad respecto al ánodo?
c) Si, además, el flujo radiante incidente es de 1 pW, ¿cuál es la corriente de salida?

17

Sensores de efecto Hall

17.1 El efecto Hall
17.2 Sensores Hall de salida lineal
17.3 Sensores Hall de salida digital
17.4 Modos de operación
17.5 Dispositivos de medida basados en el efecto Hall
17.6 Ejemplos de aplicaciones

El "efecto Hall" fue descubierto en 1879 por Edwin Herbert Hall mientras trabajaba en su tesis doctoral en la Universidad John Hopkins de Baltimore (USA). Básicamente, Hall descubrió que cuando un conductor, por el que circula una corriente eléctrica, se sitúa en un campo magnético, se genera una tensión perpendicular a ambos. Sin duda, no podía imaginarse en aquellos momentos la multitud de aplicaciones que de tal descubrimiento se iban a derivar, sobre todo si tenemos en cuenta lo poco que en esos momentos se conocía de la electricidad (el electrón no fue identificado hasta 10 años después). Aunque el descubrimiento realizado por Hall fue bien acogido, la primera aplicación práctica no vio la luz hasta la aparición en 1950, de los materiales semiconductores.

17.1 El efecto Hall

La figura 17.1 muestra una fina lámina de material metálico o semiconductor (elemento Hall) de espesor **d** a través de la cual circula una corriente eléctrica **I**. En ausencia de campo magnético, la distribución de la corriente es uniforme y no se tiene una diferencia de potencial en los terminales de salida. En cambio, cuando se encuentra presente un campo magnético de densidad **B**, los electrones se ven sometidos a una fuerza de Lorentz resultando una diferencia de potencial en la salida (tensión Hall, V_H), cuyo valor es directamente proporcional a la corriente eléctrica y al campo magnético e inversamente proporcional al espesor de la lámina:

$$V_H = R_H \cdot I \cdot B / d \qquad (17.1)$$

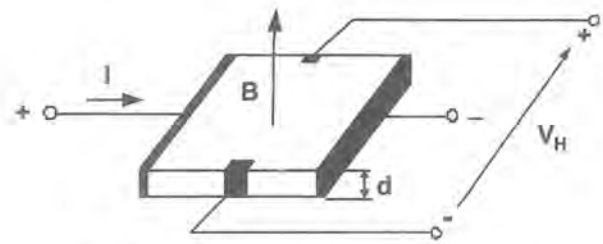


Figura 17.1. Principio básico del efecto Hall.

donde R_H es la constante de Hall, dada por:

$$R_H = -\frac{1}{ne} \text{ (semiconductor tipo n)}; \quad R_H = +\frac{1}{pe} \text{ (semiconductor tipo p)} \tag{17.2}$$

siendo n la densidad de electrones, p la densidad de huecos y e la carga del electrón.

Lógicamente, si se invierte el sentido del campo magnético o el de la corriente, cambia la polaridad de V_H. Si en la expresión (17.1) la corriente es constante, la tensión Hall será proporcional a la densidad del campo. Se tiene así una primera aplicación práctica de estos dispositivos: la medida de un campo magnético. Pero, ¿es posible medir otras magnitudes? En efecto, para ello sólo es necesario convertir dicha magnitud en un campo magnético susceptible de poder ser medido con un sensor de efecto Hall (figura 17.2). La segunda cuestión que se plantea es cómo generar el campo magnético. Esto puede realizarse a partir de una corriente eléctrica o bien empleando un imán permanente.

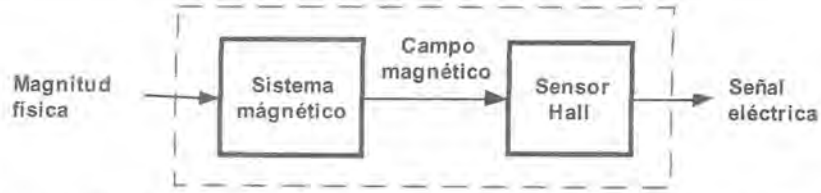


Figura 17.2. Elementos de un dispositivo de medida basado en el efecto Hall.

En la fabricación de elementos Hall se emplean materiales semiconductores como el antimoniuro de indio (InSb), el arseniuro de galio (GaAs) o el arseniuro de indio (InAs). Este último es el material más adecuado para muchas aplicaciones pues tiene un coeficiente de temperatura muy bajo (0,1%/°C) y su resistencia también es baja, contribuyendo a un menor calentamiento y ruido de origen térmico.

Los dispositivos basados en el efecto Hall tienen, como veremos, un sinfín de aplicaciones. Ello es debido a las buenas características que presentan:

- Son dispositivos de estado sólido.
- Larga vida de funcionamiento (millones de operaciones de forma continuada).
- Frecuencia de funcionamiento relativamente alta (hasta 100 kHz o más).
- No tienen partes móviles.
- Compatibilidad con otros circuitos electrónicos.
- Margen de temperatura amplio.
- Buena repetibilidad.

Atendiendo al tipo de salida, los sensores Hall pueden ser:

- Sensores Hall de salida lineal o analógicos.
- Sensores Hall de salida digital.

17.2 Sensores Hall de salida lineal

La figura 17.3 muestra la estructura básica de un sensor Hall de salida lineal. La tensión de salida del elemento Hall requiere un acondicionamiento que consiste básicamente en una etapa de amplificación diferencial, dado que la tensión Hall es de muy bajo nivel (≈30 μV/G), y una compensación de temperatura. También suelen integrar un regulador de tensión a fin de poder trabajar con amplio margen de tensiones de alimentación y mantener la corriente constante de forma que la tensión de salida refleje sólo la intensidad del campo magnético. Para permitir una mejor interfaz con otros dispositivos, a la salida del amplificador diferencial se añade un transistor en emisor abierto, colector abierto o *push-pull*.

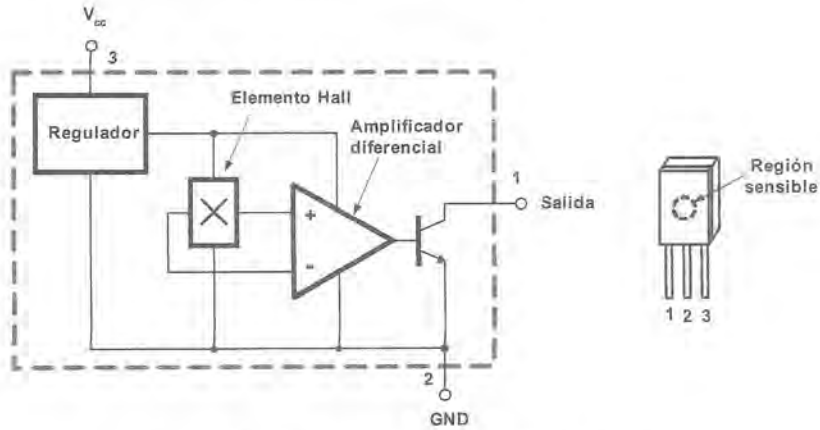


Figura 17.3. Estructura básica de un sensor lineal.

Según sea la polaridad del campo magnético, la tensión de salida del amplificador podrá ser positiva o negativa, lo cual requiere disponer, en principio, de fuentes de alimentación tanto positivas como negativas. Para evitar esta situación, en el amplificador diferencial se incorpora una tensión de polarización, de forma que con un campo magnético nulo se tenga en la salida una tensión positiva. Esta tensión se conoce, dependiendo del fabricante, como ***null offset*** o ***quiescent output voltage***.

La figura 17.4 muestra las características de transferencia de un sensor Hall lineal [17.1]. Además del *null offset*, sobre dichas características se definen otros dos parámetros:

- **Sensibilidad** (mV/G). Es la pendiente de la característica de transferencia. En la mayoría de los sensores Hall analógicos la sensibilidad y el *offset* son proporcionales a la tensión de alimentación. Se dice que son sensores ratiométricos.
- **Alcance** (*span*). Define el margen de salida del sensor, es decir: V_o @ $B_{máx}$ – V_o @ B_{min}.

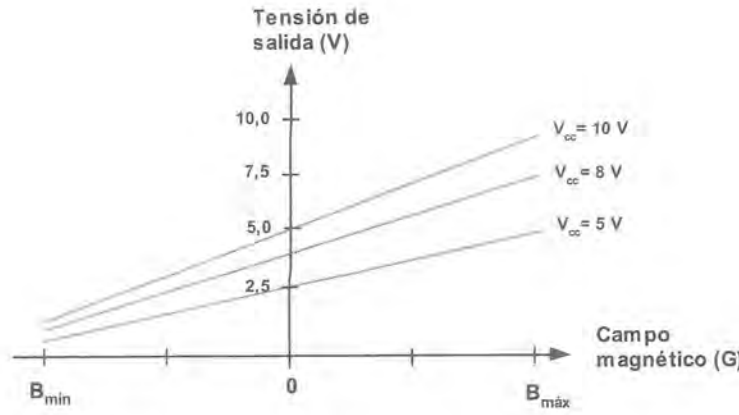


Figura 17.4. Características de transferencia de un sensor Hall lineal.

EJERCICIO 17.1

Un sensor Hall de salida lineal tiene una función de transferencia como la mostrada en la figura 17.4, en la que la densidad de flujo magnético máximo es de ±900 G[1]*. Si se alimenta con +5 V, determine el valor del null offset, la sensibilidad y el span. ¿Es su funcionamiento rail-to-rail?*

SOLUCIÓN:

De la figura 17.4 tenemos que la tensión de salida con campo magnético nulo (*null offset*) es de 2,5 V. La sensibilidad es la pendiente de la recta correspondiente a una tensión de alimentación de 5 V, es decir, vale: (5–2,5)/900 = 2,7 mV/G. El *span* es de aproximadamente 5 V.

Es *rail-to-rail* ya que la salida varía entre aproximadamente 0 (generalmente 0,2 V) y 5 V (V_{cc} – 0,2 V).

Especificaciones eléctricas

A fin de familiarizarnos con las especificaciones eléctricas típicas de los sensores Hall de salida lineal la tabla 17.1 muestra las características eléctricas del sensor Hall lineal 3503 de Allegro MicroSystems [17.2].

Características eléctricas T_A = +25°C y V_{CC} = 5V					
Característica	**Símbolo**	**Mín.**	**Típico**	**Máx.**	**Unidades**
Tensión de alimentación	V_{CC}	4,5	–	6,0	V
Corriente de alimentación	I_{CC}	–	9,0	14	mA
Tensión de salida para B = 0	V_{OUT}	2,25	2,50	2,75	V
Sensibilidad	ΔV_{OUT}	0,75	1,30	1,72	mV/G
Ancho de banda (–3dB)	BW	–	23	–	kHz
Resistencia de salida	R_{OUT}	–	50	–	Ω
Histéresis	B_{hys}	30	52	–	G

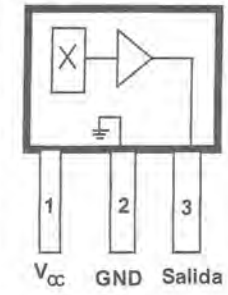


Tabla 17.1. Especificaciones del sensor Hall lineal 3503 de Allegro MicroSystems.

[1] En el Apéndice A17.1 se recogen las magnitudes magnéticas básicas, así como sus unidades de medida.

La densidad de flujo magnético no tiene un valor máximo, de forma tal que un valor excesivo del mismo en ningún caso dañará el sensor.

Circuitos de interfaz

La salida típica de estos sensores suele ser una configuración en emisor abierto, como se muestra en la figura 17.5. El transistor de salida suministra una corriente a la carga de forma que la tensión de salida, V_o es proporcional al campo magnético que se tiene en la superficie sensible del sensor.

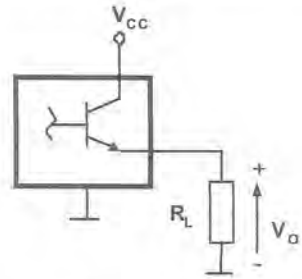


Figura 17.5. Etapa de salida de un sensor Hall analógico.

Los sensores Hall pueden conectarse con facilidad a componentes estándar. Son bastante frecuentes las aplicaciones donde la salida se lleva a un circuito comparador o a un amplificador. La tabla 17.2 muestra algunos ejemplos de circuitos de interfaz básicos [17.7]. Empleando un circuito comparador con alimentación única (por ejemplo, el LM339 o equivalente) se puede obtener un interruptor digital con punto de funcionamiento ajustable. La resistencia R_H proporciona la histéresis a los circuitos comparadores. La exactitud de la medida depende de la tolerancia y coeficiente de temperatura de las resistencias, sin embargo, la resistencia R_L no es crítica.

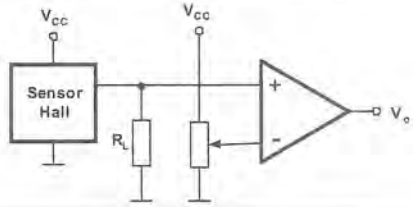
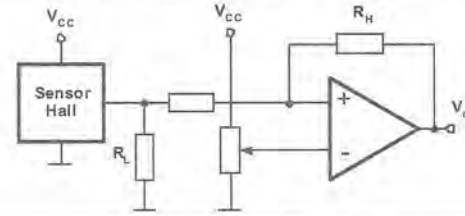
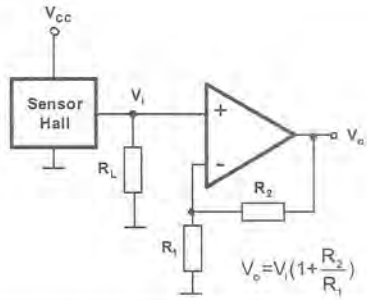
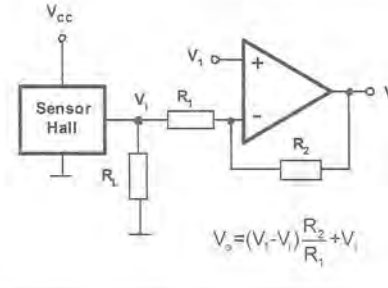
Comparador no inversor	Comparador no inversor con histéresis
	
Amplificador no inversor	**Amplificador inversor**
	

Tabla 17.2. Circuitos de interfaz básicos para sensores Hall lineales.

17.3 Sensores Hall de salida digital

Si, a la estructura básica de la figura 17.3, añadimos un comparador *Schmitt-Trigger* se obtiene un sensor Hall de salida digital, cuyo diagrama de bloques se muestra en la figura 17.6, y su característica de transferencia en la figura 17.7.

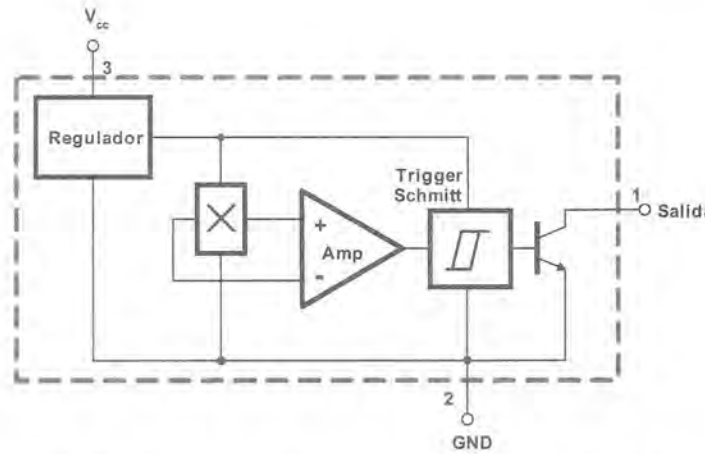


Figura 17.6. Diagrama de bloques funcional de un sensor Hall digital.

En ausencia de campo magnético, el transistor de salida está cortado (estado *OFF*). Cuando el campo magnético, perpendicular a la superficie del chip, está por encima de un determinado valor umbral, punto

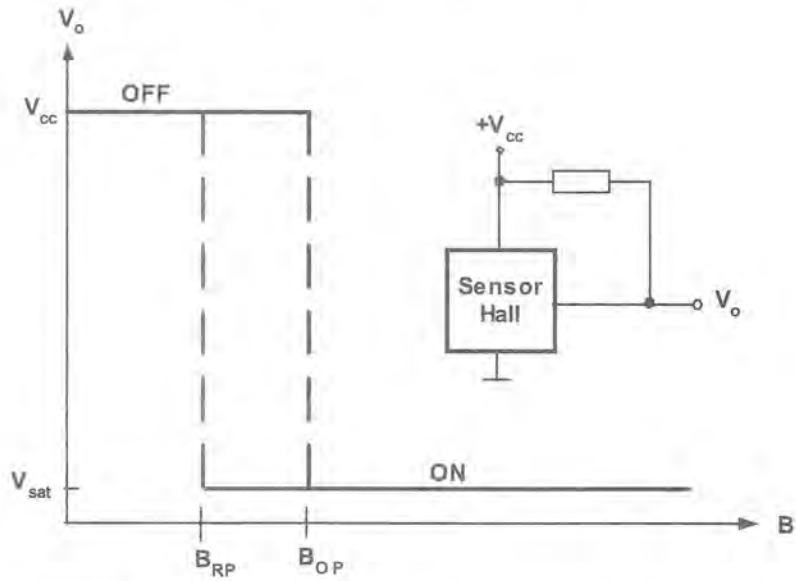


Figura 17.7. Característica de transferencia de un sensor Hall de salida digital.

B_{OP} de la característica (*operating point*), el transistor de salida conmuta a saturación (estado *ON*), pudiendo conducir corriente. Si ahora vamos reduciendo el campo magnético, para un valor inferior al punto B_{RP} de la característica (*release point*), el transistor se cortará. Para evitar posibles conmutaciones erróneas que se puedan producir en la salida como consecuencia de pequeñas variaciones en la densidad de flujo magnético, el circuito tiene una cierta histéresis (la diferencia entre los puntos B_{OP} y B_{RP}).

Dado que las características varían con la temperatura y de un sensor a otro, los fabricantes especifican unos valores máximos y mínimos de densidad de flujo magnético para la conmutación de la salida del sensor Hall, B_{OP} y B_{RP}, respectivamente. Según sea el valor de B_{OP} y B_{RP}, los sensores de salida digitales pueden ser:

- **Unipolares**, cuando tanto el valor máximo de B_{OP} como el valor mínimo de B_{RP} es positivo.
- **Bipolares**, cuando el valor máximo de B_{OP} es positivo y el mínimo de B_{RP} es negativo. Si el estado de la salida se mantiene en ausencia de campo magnético, el sensor se denomina ***latch* bipolar.**

La tabla 17.3 muestra las especificaciones eléctricas y magnéticas típicas del interruptor Hall bipolar 3132 de Allegro MicroSystems [17.2].

Características eléctricas a $T_A = 25°C$					
Característica	**Símbolo**	**Mín.**	**Típico**	**Máx.**	**Unidades**
Tensión de alimentación	V_{CC}	4,5	–	24	V
Tensión de salida en *"on"*	$V_{OUT(SAT)}$	–	145	400	mV
Corriente de salida en *"off"*	I_{OFF}	–	<1,0	10	μA
Corriente de alimentación	I_{CC}	–	4,3	9,0	mA
Tiempo de subida	t_r	–	0,04	2,0	μs
Tiempo de bajada	t_f	–	0,18	2,0	μs
Características magnéticas					
Punto de activación (OP)	B_{OP}	–	32	95	G
Punto de desactivación (RP)	B_{RP}	–95	–20	–	G
Histéresis	B_{hys}	30	52	–	G

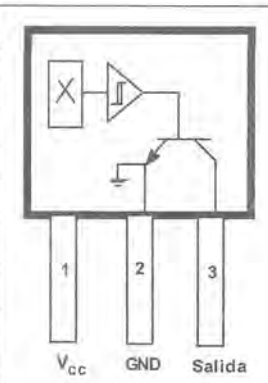


Tabla 17.3. Especificaciones del sensor bipolar 3122 de Allegro MicroSystems.

Circuitos de interfaz

La salida típica de estos sensores Hall es en colector abierto, por lo que se requiere una resistencia entre la salida y la alimentación (resistencia de *pull-up*) que garantice una salida a nivel alto cuando el transistor esté cortado (figura 17.8). Cuando el sensor se activa por la presencia de un campo magnético, el transistor estará *ON* absorbiendo la corriente I_L que circula por la carga. La tabla 17.4 muestra diversos ejemplos de interfaz de sensores Hall de salida digital.

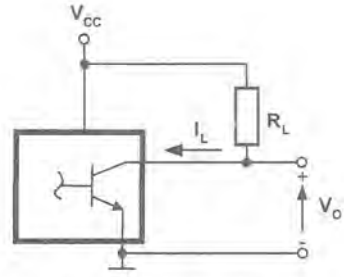


Figura 17.8. Salida en colector abierto de un sensor Hall de salida digital.

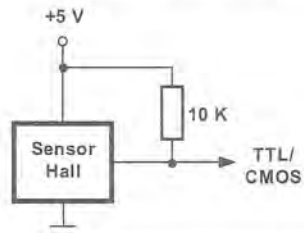
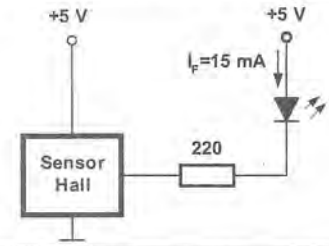
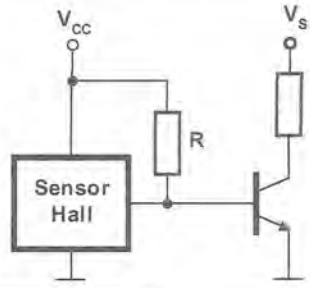
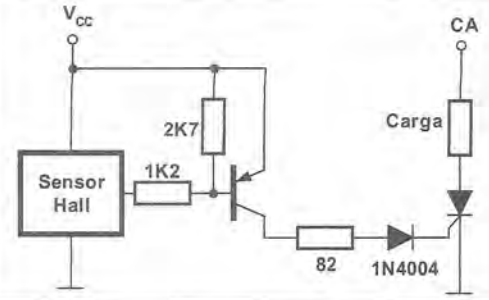
Interfaz con lógica TTL y CMOS	Interfaz con un LED
	
Interfaz con un transistor	**Interfaz con un tiristor**
	

Tabla 17.4. Circuitos de interfaz de sensores Hall de salida digital.

17.4 Modos de operación

Cuando se desea medir una magnitud física (posición, velocidad, temperatura, etc.) mediante un sensor de efecto Hall, es necesario realizar una conversión previa de la misma a un campo magnético proporcional. Esto requiere que el imán permanente experimente algún tipo de movimiento característico (modo de operación). Estos modos pueden ser **unipolares** (si el campo magnético es unipolar) o **bipolares** (campo bipolar). Los modos unipolares pueden ser:

- **Modo frontal (*Head-on mode*)**. La dirección de movimiento del imán es perpendicular a la cara activa del sensor, por lo que las líneas de flujo magnético atraviesan dicha cara activa (figura 17.9). La densidad del campo magnético es inversamente proporcional a la distancia de separación D. Este modo es simple, funciona bien y es relativamente insensible al movimiento lateral.

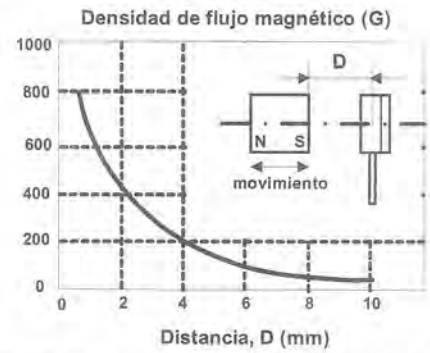


Figura 17.9. Modo de operación unipolar frontal.

– **Modo lateral (*Slide-by mode*).** La dirección de movimiento del imán es paralela a la cara activa del sensor (figura 17.10). La curva en este caso tiene forma de campana. Lógicamente, se tendrá una curva para cada distancia de separación (*gap*) entre el plano del imán y el del sensor.

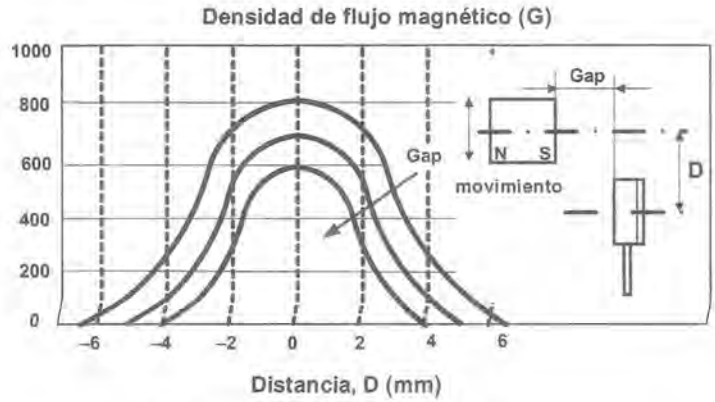


Figura 17.10. Modo de operación unipolar lateral.

Los modos bipolares pueden adoptar también diferentes configuraciones [17.1], [17.6]. La figura 17.11 muestra el ejemplo de un modo bipolar lateral en el que se tiene un imán en anillo con dos pares de polos. Dependiendo de la aplicación se puede disponer de diferente número de pares de polos. El movimiento de rotación del imán da lugar a una densidad de campo magnético con una forma de onda senoidal. Si se emplea con un sensor Hall de salida digital, en cada revolución se obtendrá un pulso de salida por cada par de polos. Cuantos más pares de polos, la amplitud de la onda es menor.

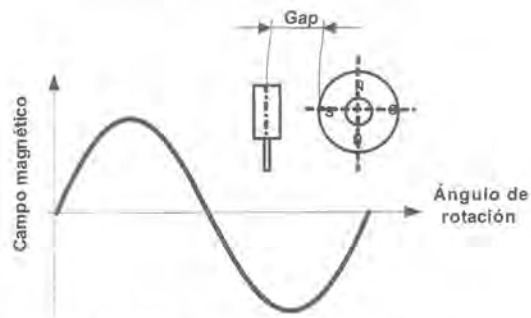


Figura 17.11. Modo bipolar lateral (imán en anillo).

Con objeto de incrementar la densidad del flujo magnético, es frecuente el uso de **piezas polares** o **concentradores de flujo**. Si se añaden a un sistema magnético ofrecen un camino de menor resistencia a las líneas de flujo. Cuando se colocan enfrentadas a la cara de un imán (figura 17.12a), la densidad de flujo en el aire entre ambas se incrementa. La figura 17.12b muestra la variación del campo magnético en el modo *unipolar frontal. El hecho de añadir la pieza polar aporta tres beneficios: un incremento en la densidad de* flujo magnético, la distancia de actuación del sensor es mayor (para un nivel de flujo constante) y se puede emplear un imán con una intensidad de campo menor, por tanto, más barato. Estos concentradores suelen estar integrados en el sensor.

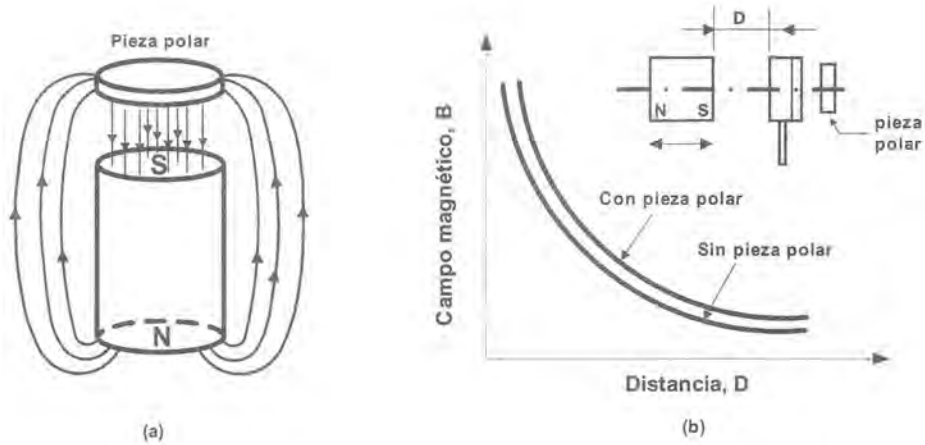


Figura 17.12. Efecto de una pieza polar (concentrador de flujo).

17.5 Dispositivos de medida basados en el efecto Hall

En este punto se presentarán tres dispositivos de medida basados en el efecto Hall que se pueden encontrar comercialmente. Estos dispositivos integran en un mismo encapsulado el elemento Hall y el sistema magnético y son la base de multitud de aplicaciones.

Interruptor Hall de ranura

Detectan la presencia (o ausencia) de una lámina de metal ferromagnético dentada cuando pasa por el hueco existente entre un imán y un sensor Hall de salida digital, produciendo una salida digital sin rebotes.

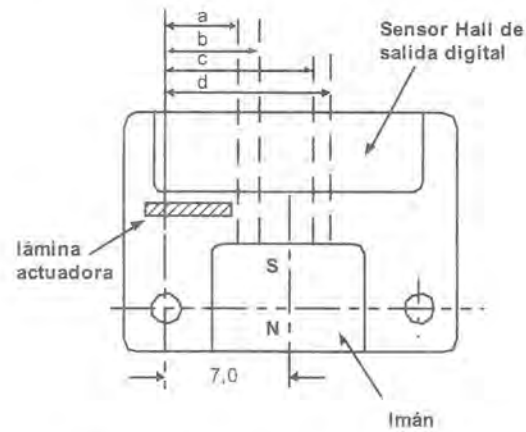


Figura 17.13. Interruptor Hall de ranura.

La figura 17.13 muestra los detalles de actuación de este dispositivo Hall [17.9]. Inicialmente el sensor se encuentra en el estado *ON*. Si la lámina se mueve de izquierda a derecha, cuando alcance el punto **b** la salida del sensor, pasará al estado *OFF* y se mantendrá así hasta que la lámina llegue al punto **d** en el que volverá a conmutar al estado *ON*. Si la lámina se mueve de derecha a izquierda, el sensor conmutará a *OFF* en el punto **c** y a *ON* en el punto **a**. Para detectar un movimiento de rotación se tienen láminas dentadas en configuración circular. Estos dispositivos se utilizan en aplicaciones de recuento, posicionamiento, automoción, detección de parada, tacómetros y pueden ser usados en ambientes con polvo o con fuerte iluminación donde un detector óptico no sería apropiado.

Sensor Hall de corriente

Un sensor de corriente con salida analógica puede realizarse mediante un núcleo de ferrita y un sensor Hall lineal montados en un encapsulado plástico. La corriente que circula por un conductor crea un campo magnético. Este campo magnético es llevado por el núcleo de ferrita a un sensor Hall, que convierte dicho campo en una tensión de salida proporcional a la corriente en el conductor.

La medida de corriente puede realizarse en bucle abierto o en bucle cerrado. En la **medida en bucle abierto** (figura 17.14), la corriente a medir I, circula por el conductor que atraviesa el interior del núcleo magnético, produciendo un campo magnético proporcional a la magnitud de la corriente. El núcleo concentra el campo magnético en el sensor Hall, el cual genera una tensión proporcional a dicho campo y, por tanto, a la corriente a medir. La tensión Hall de salida es de muy pequeño valor (μV) por lo que se requiere una etapa de amplificación. La linealidad del sensor está determinada por las características del núcleo magnético y del generador Hall. Las principales ventajas de esta técnica son su bajo consumo, pequeño tamaño y poco peso, por lo que se emplea en circuitos alimentados desde baterías.

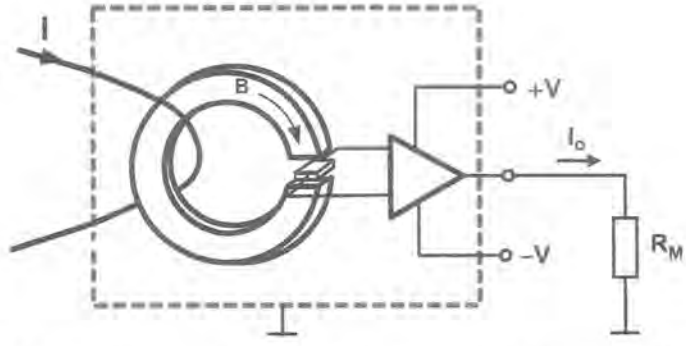


Figura 17.14. Medida de corriente en bucle abierto mediante un dispositivo Hall.

En la **medida en bucle cerrado** (figura 17.15), la tensión de salida del sensor Hall es previamente amplificada y después excita una etapa en *push-pull* cuya corriente de salida circula por una bobina secundaria arrollada en el núcleo magnético. El campo magnético creado en la bobina secundaria cancela el flujo del primario, manteniendo un nivel de flujo en el núcleo próximo a cero. Se cumple que: $I_P \cdot N_P = I_S \cdot N_S$. Generalmente, $N_P = 1$, con lo que la corriente en el secundario queda dividida por N_S. Por ejemplo, un sensor con $N_S = 1000$ (relación 1:1000) proporciona una salida de 1 mA/A. El valor de la corriente I_S se puede obtener midiendo la caída de tensión en una resistencia R_M. Los fabricantes suelen especificar un valor mínimo de R_M que garantice la adecuada disipación de potencia del circuito y un valor máximo de R_M que evite su saturación.

La medida en bucle cerrado tiene significativas mejoras respecto al circuito en bucle abierto: al trabajar con un flujo magnético aproximadamente nulo permite eliminar la influencia de las no linealidades del nú-

cleo magnético y reducir los efectos de la sensibilidad a la temperatura en el elemento Hall. En el capítulo 19 se tratará la aplicación de los sensores Hall para la medida de otras magnitudes eléctricas.

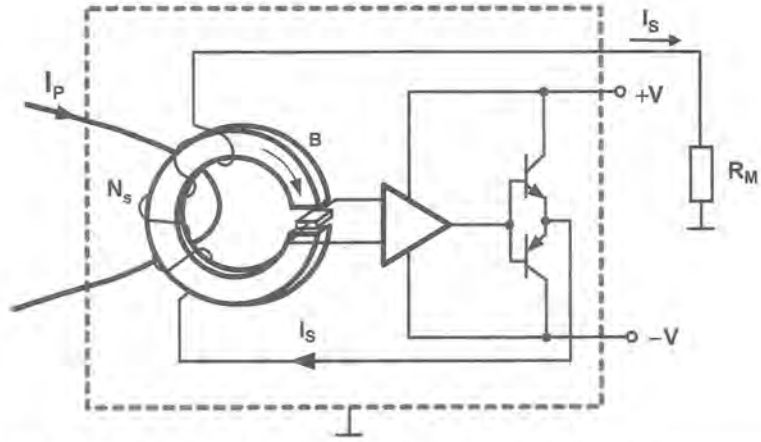


Figura 17.15. Medida de corriente en bucle cerrado mediante un dispositivo Hall.

Sensor de dientes de engranaje

Se trata de un dispositivo encapsulado en una carcasa tipo sonda, que permite detectar con gran precisión el movimiento de un objeto de material ferromagnético. La figura 17.16 muestra su aplicación como sensor de dientes de engranaje. Cuando el diente del engranaje pasa por la cara del sensor, detecta un cambio en el nivel del flujo y traslada este cambio a la tensión de salida, que conmuta entre la tensión de alimentación y la tensión de saturación del transistor de salida.

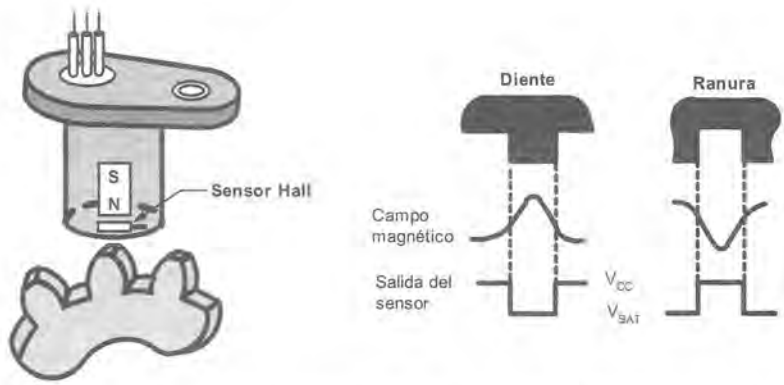


Figura 17.16. Lector de efecto Hall empleado para detectar los dientes de un engranaje.

17.6 Ejemplos de aplicaciones

Las buenas características de los sensores de efecto Hall, unidas a la bajada de precio que estos dispositivos han experimentado han dado lugar a un importante crecimiento en el número de aplicaciones, como se pone

de manifiesto en la tabla 17.5, de hecho, aplicaciones donde tradicionalmente se utilizaban sensores mecánicos y ópticos se realizan actualmente con sensores Hall. Como se puede observar los sensores de efecto Hall tienen un gran número de aplicaciones en el sector del automóvil.

Aplicaciones en el automóvil		Aplicaciones industriales	
Sensor Hall digital	**Sensor Hall lineal**	**Sensor Hall digital**	**Sensor Hall lineal**
• Cigüeñal • Árbol de levas • Encendido • Motor de arranque • ABS • Ajuste de asientos • Boqueo de cerradura • Control tracción • Elevalunas • Techo corredizo • Ajuste de espejos • Ajuste de asientos • Cinturones seguridad	• Control suspensión • Medida de corriente • Nivel de líquido • Gestión de la batería • Dirección asistida • Frenos • Ajuste de los faros	• Motores DC sin escobillas • Ventiladores • Medida de velocidad • Interruptor sin contactos • Nivel de líquidos • Flujo de líquidos • Interruptor de presión • Interruptor de rotación	• Medida de corriente • Medida de posición • Medida de distancia • Media de presión • Medida de fuerza • Medida de campo magnético • Medida de par

Tabla 17.5. Principales aplicaciones de los sensores de efecto Hall.

Medida de variables físicas

La medida de presión y temperatura puede realizarse con la ayuda del esquema de la figura 17.17a. Cuando se expande o contrae el fuelle se mueve el imán hacia las proximidades del sensor Hall, proporcionando una tensión de salida proporcional a la presión. Para la medida de la temperatura, un gas con una constante térmica conocida está encerrado dentro del fuelle. Cuando la cámara se calienta, el gas se expande dando el sensor una tensión proporcional a la temperatura.

La figura 17.17b muestra la medida de nivel de un líquido en un tanque de material no ferromagnético. El imán está unido al flotador de forma que cuando el nivel del tanque varía, el sensor Hall detectará estas variaciones y dará en su salida una tensión proporcional. De esta forma, se evitan conexiones eléctricas en el interior del tanque.

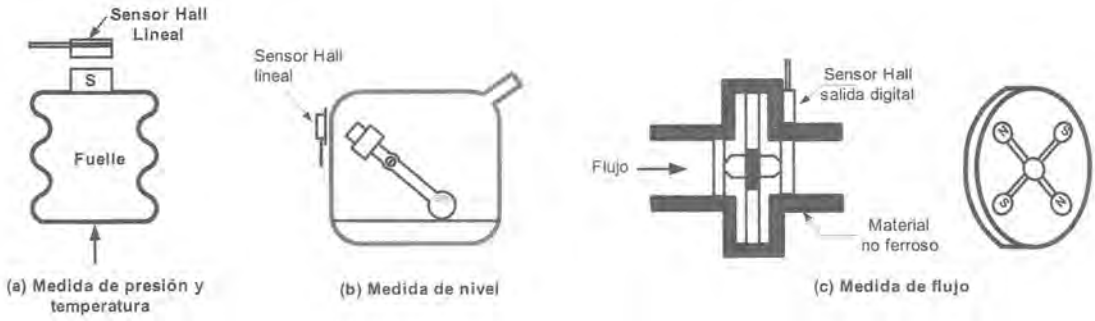


Figura 17.17. Medida de variables físicas con sensores Hall.

La medida de flujo de agua puede realizarse mediante el esquema de principio de la figura 17.17c. El agua mueve unas aspas en las que se tienen dos pares de polos. Cuando estos imanes pasan por el sensor

Hall, este genera dos pulsos de tensión por revolución. De esta forma, se puede conocer la cantidad de agua que circula, haciendo un uso más eficiente de la misma.

Sistema de control analógico de posición

La figura 17.18 muestra el control de la posición de un sistema mecánico [17.6]. Dicha posición es controlada mediante el potenciómetro, el cual se ajusta para obtener una determinada salida. Si se produce una variación en la posición, el imán (solidario a la cabeza de movimiento) variará su posición y, por tanto, la tensión de salida del sensor Hall, con lo que varía la salida del servoamplificador que gobierna el servomotor, a fin de corregir la variación de la posición. Este principio se utiliza en los sistemas de aire acondicionado, ventilación, así como en el control de procesos.

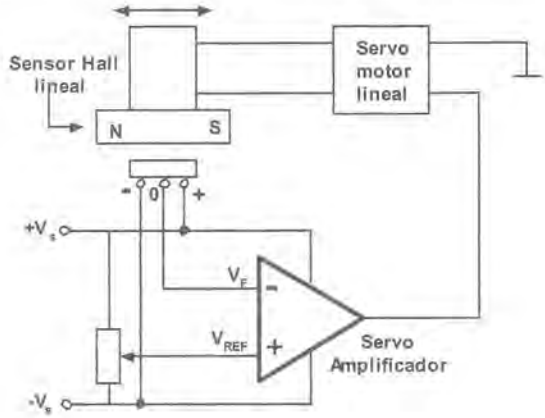


Figura 17.18. Control analógico de posición.

Motor CC sin escobillas

La figura 17.19 muestra los elementos esenciales de un típico motor CC sin escobillas [17.6]. Se emplean tres sensores Hall bipolares de salida digital, activados por los imanes permanentes montados en el eje del rotor. Los sensores Hall miden la posición angular del eje, proporcionando esta información a un circuito lógico que codifica esta información y controla la activación de los interruptores del circuito de gobierno. Cuando el polo Sur pasa delante de la cara activa del sensor la salida de este cambia de estado. La figura 17.19 muestra también las formas de onda que se tienen en la salida de los tres sensores Hall, correspondientes a una separación entre sensores de 30 grados eléctricos.

La conmutación electrónica ofrece una gran flexibilidad de interfaz con sistemas de control electrónico. Por otro lado, la supresión de las escobillas elimina los problemas de ruido y rozamiento.

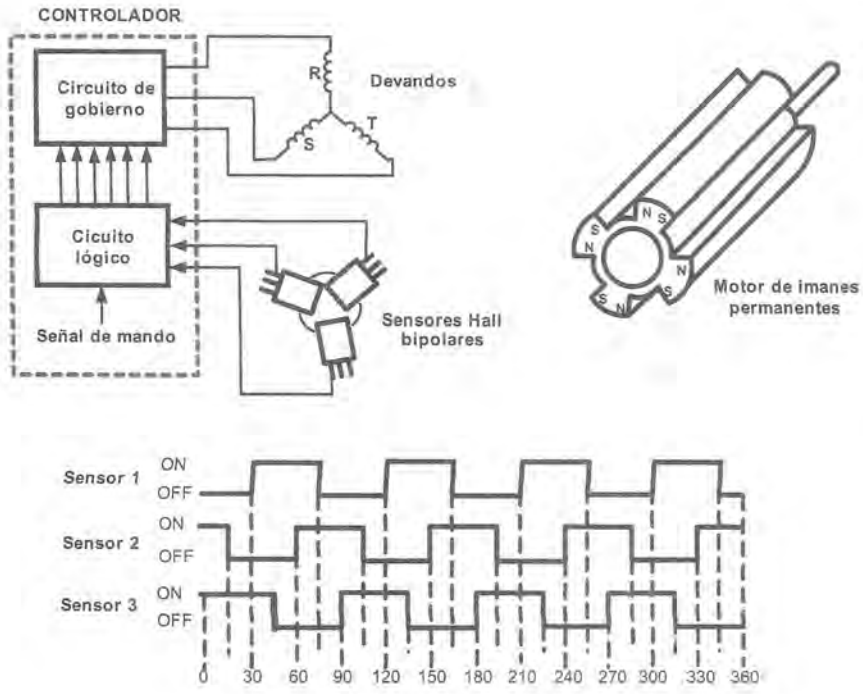


Figura 17.19

Resumen

- *En este capítulo se presentaron los fundamentos y aplicaciones de los sensores basados en el efecto* Hall. Se trata de dispositivos de estado sólido que se caracterizan por: no tener partes móviles, frecuencia de funcionamiento relativamente alta, margen de temperatura amplio y ser compatibles con otros circuitos electrónicos, etc.
- Los sensores de efecto Hall pueden ser de salida lineal, en los que la salida varía linealmente con el campo magnético, o de salida digital, en los que la salida sólo puede conmutar entre dos niveles de tensión. Para cada uno de estos tipos se expusieron sus características eléctricas más significativas, así como los circuitos de interfaz más característicos.
- Cuando se desea medir una magnitud física mediante un sensor de efecto Hall se requiere realizar una conversión previa de la magnitud a medir en un campo magnético proporcional. Esto supone que el imán permanente ha de experimentar algún tipo de movimiento característico (modo de operación). Según la polaridad del campo magnético los modos de operación pueden ser unipolares o bipolares.
- Con objeto de incrementar la densidad de flujo magnético, es frecuente el empleo de piezas polares o concentradores de flujo. Esto permite una mayor distancia de actuación, así como la posibilidad de emplear un imán con una intensidad de campo menor y, por tanto, más barato.
- Se presentaron tres dispositivos comerciales basados en el efecto Hall: interruptor de ranura, sensor Hall de corriente y sensor de dientes de engranaje. Estos dispositivos integran en un mismo encapsulado el elemento Hall y el sistema magnético, siendo la base de multitud de aplicaciones.

- Finalmente se expusieron algunas de las aplicaciones más significativas de los sensores de efecto Hall. Los dispositivos Hall forman parte de una gran variedad de equipos: máquinas herramientas, equipos médicos, automóviles, computadores, etc. El número de aplicaciones sólo está limitado por la capacidad de imaginación del diseñador.

APÉNDICE A17.1

Magnitud	Símbolo	SI	CGS	Relación
Densidad del flujo magnético	**B**	Tesla, T	Gauss, G	$1\ G = 10^{-4}\ T$
Flujo magnético	$\Phi = \int BdS$	Weber, Wb	Maxwell, M	$1\ M = 10^{-8}\ Wb$
Intensidad de campo magnético	**H**	A/m	Oersted, Oe	$1\ Oe = 10^3/4\pi\ A/m$
Fuerza magnetomotriz	**F**	Amperio, A	Gilbert, Gb	1 Gb = 1 Oe x cm
Reluctancia	$\mathfrak{R}$	1/henrio, H^{-1}	Gilbert/Maxell	$1\ Gb/M = 10^9/4\pi\ H^{-1}$

Tabla 17.6. Magnitudes magnéticas básicas.

Características magnéticas de los imanes permanentes

- **Inducción residual (B_R)**: es la densidad de flujo disponible en el material magnético cuando el campo magnético es nulo.
- **Fuerza coercitiva (H_C)**: es el campo magnético que debe ser aplicado en dirección opuesta a la inducción residual para anular la densidad de flujo.
- **Producto BH máximo (BH_{max})**: este parámetro se utiliza para comparar un material magnético con otro. Cuando mayor sea este producto más "resistente" es el material magnético a las fuerzas de desmagnetización.

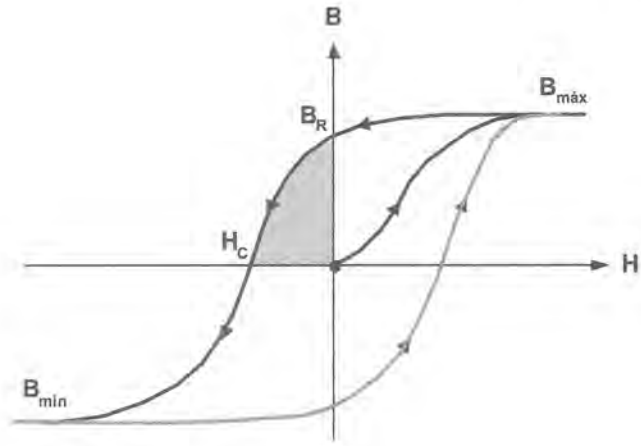


Figura 17.20. Ciclo de histéresis.

Material	B_R (G)	H_C (Oe)	$BH_{máx}$ (G-Oe)	Coste	Coef. de Temp. (%/°C)
Tierras raras-cobalto	$8{,}1 \cdot 10^3$	$7{,}9 \cdot 10^3$	$16 \cdot 10^6$	Muy alto	–0,05
Alnico 1 (Ni, Al, Co, Fe)	$7{,}1 \cdot 10^3$	440	$1{,}4 \cdot 10^6$	Medio	–0,02
Alnico 5 (Co, Ni, Al, Cu, Fe)	$13{,}1 \cdot 10^3$	640	$6 \cdot 10^6$	Medio-alto	–0,02
Alnico 12 (Co, Ni, Ti, Al, Fe)	$5{,}8 \cdot 10^3$	950	$1{,}6 \cdot 10^6$	Alto	–0,02
Cerámico 1 (Ba, Sr, Fe)	$2{,}2 \cdot 10^3$	$1{,}8 \cdot 10^3$	$1{,}0 \cdot 10^6$	Bajo	–0,2
Cunife (Cu, Fe, Ni)	$5{,}5 \cdot 10^3$	$0{,}53 \cdot 10^3$	$1{,}4 \cdot 10^6$	Medio	—
Fe-Cr	$13{,}5 \cdot 10^3$	$0{,}60 \cdot 10^3$	$5{,}25 \cdot 10^6$	Medio	—
Rubber (Ba, Sr, Fe)	$2{,}3 \cdot 10^3$	$1{,}8 \cdot 10^3$	$1{,}1 \cdot 10^6$	Bajo	–0,5
Neodimio (Nd, Fe, B)	$11{,}7 \cdot 10^3$	$6{,}5 \cdot 10^3$	$15 \cdot 10^6$	Medio-Alto	–0,157

Tabla 17.7. Propiedades de materiales magnéticos para imanes permanentes [17.1].

Problemas propuestos

17.1 Determine la tensión Hall que se obtiene en una placa de cobre de 1 mm de espesor situada en un campo magnético de densidad 1 T y por la que circula una corriente de 1 A. Repita el problema en el caso de tratarse de silicio tipo n. El número de portadores de carga por unidad de volumen es 10^{29} electrones/m^3 en el caso del cobre y 10^{22} para el silicio tipo n. La carga del electrón es $1{,}6 \cdot 10^{-19}$C.

17.2 Un interruptor Hall a temperatura ambiente tiene valores máximo y mínimo de activación y de desactivación (puntos OP y RP) de ±300 G. El imán está hecho de un material, cuyo coeficiente de temperatura es –0,2%/°C. Determine el valor de la densidad de campo magnético necesario si se desea que pueda funcionar adecuadamente en el margen de temperatura entre –20 y +85°C.

17.3 Un sensor Hall de salida digital está localizado en una posición remota, por lo que se realiza un *bucle de corriente a dos hilos como se muestra en la figura 17.21. Cuando el sensor no actúa la co*rriente en el bucle es básicamente igual a la corriente de alimentación del sensor. Cuando el sensor actúa, la corriente en el bucle es igual a la corriente de alimentación más la corriente que circula por el transistor. Seleccione los valores adecuados de las resistencias R_1 y R_2.

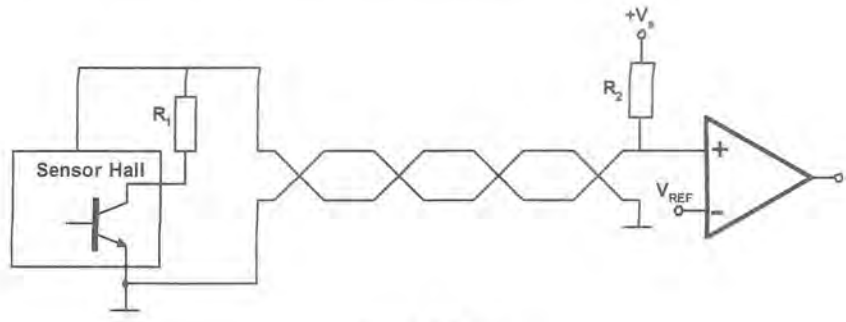


Figura 17.21

17.4 Para determinar el sentido de rotación de un imán en anillo se emplean dos dispositivos de efecto Hall de salida lineal como se muestra en la figura 17.22. Si el imán rota en el sentido de la figura, el tiempo en pasar el polo sur desde el sensor S_2 al S_1 será menor que el tiempo en dar una vuelta completa. Proponga una interfaz digital que permita determinar el sentido de rotación del imán en anillo.

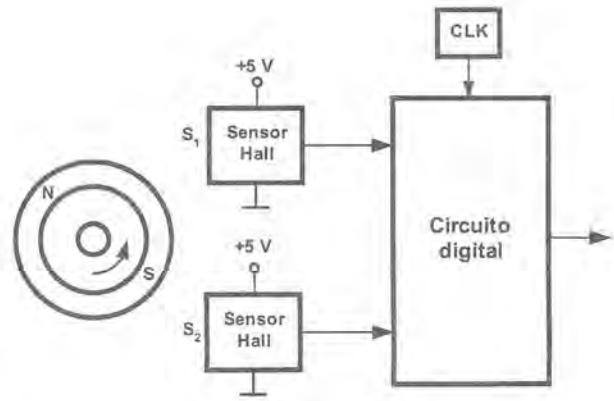


Figura 17.22

17.5 Se conectan en paralelo varios interruptores Hall con salida en colector abierto como se muestra en la figura 17.23. ¿Qué función lógica se tiene en el punto de conexión? Determine el valor mínimo y máximo de la resistencia R de *pull-up* necesaria.

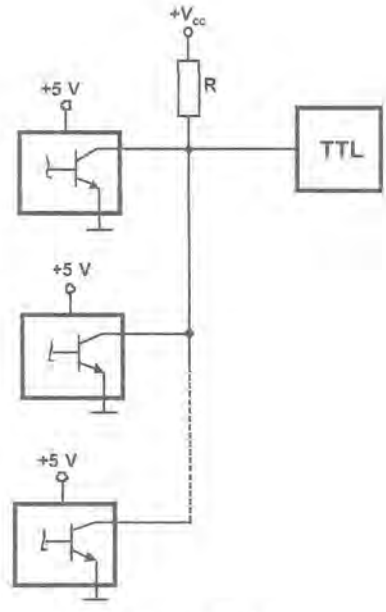


Figura 17.23

17.6 El sensor Hall de salida lineal de la figura 17.24 se emplea para detectar la ausencia de metal ferromagnético. Cuando la ranura pasa por el sensor, la salida de este cae unos 10 mV. Diseñar el circuito de acondicionamiento necesario para que circule por la carga una corriente de 0,5 A. Se desea filtrar las señales de frecuencia superior a 1 Hz.

Figura 17.24

17.7 Diseñe un circuito de tal forma que cuando el campo magnético esté comprendido entre 1500 y 2000 G la salida del mismo esté a nivel alto (*ON*). La sensibilidad del sensor Hall es de 1,3 mV/G. y la alimentación es de 5 V.

17.8 Realice la interfaz entre un sensor Hall de salida digital y un relé de 12 V que requiere una corriente de excitación de 100 mA. Sensor Hall: $V_{CE(SAT)} = 0{,}15$ V, $I_{ON} = 20$ mA.

17.9 Se desea construir un amperímetro de continua entre 0 y 5 A. Para ello se monta el sistema de medida de la figura 17.25, compuesto por un núcleo de ferrita y un sensor Hall de sensibilidad 1,3 mV/G. Determinar el número de espiras de la bobina, sabiendo que la tensión Hall es de 3,5 V *cuando circulan 5 A.* $L_a = 2{,}5$ *mm.*

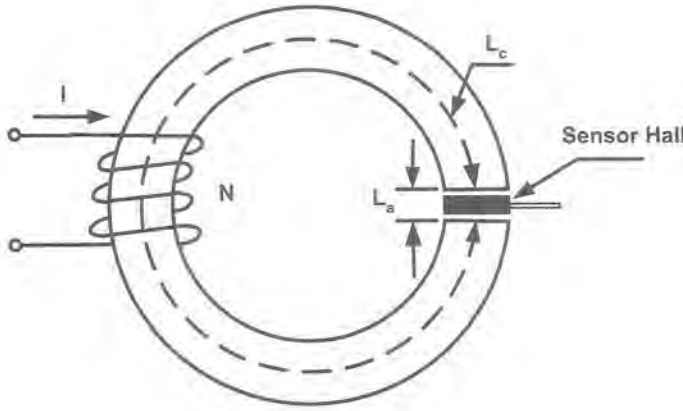


Figura 17.25

17.10 Un sensor Hall de tensión tiene las siguientes características: tensión nominal primaria $V_{PN} = 100$ V, margen de medida de tensión primaria, V_P entre −1700 V y +1700 V; corriente nominal primaria I_{PN} =10m A; relación de espiras: 10000/2000; resistencias internas del primario y del secundario (a +70°C): $R_P = 1900\ \Omega$, $R_S = 60\ \Omega$; resistencia de medida $R_{Mmáx} = 150\ \Omega$. Determinar el valor de la resistencia externa R_1 necesaria para medir una tensión $V_{PN} = 230$ V con una tensión de pico de 500 V. La figura 19.26 muestra el circuito equivalente de entrada y de salida del sensor Hall.

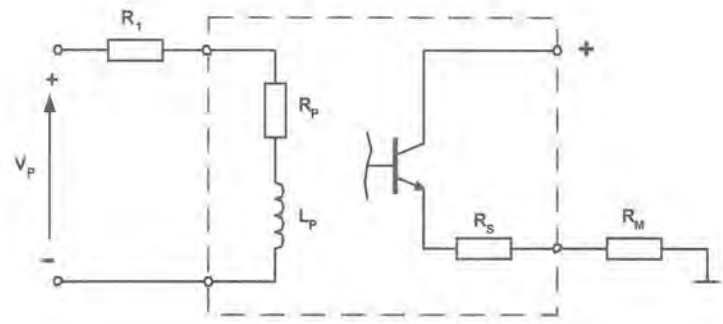


Figura 17.26

18

Otros tipos de sensores

18.1 Sensores de fibra óptica
18.2 Biosensores

18.1 Sensores de fibra óptica

La tecnología y aplicaciones de la fibra óptica han progresado muy rápidamente en los últimos años. Esto ha supuesto que si bien en un principio las aplicaciones se orientaron hacia las comunicaciones por fibra óptica, en seguida se vio la posibilidad de llevarlas al campo de los sensores. Las ventajas fundamentales que aporta la tecnología de sensores basados en la fibra óptica son:

- Mayor sensibilidad en general que los sensores tradicionales.
- Empleo en entornos hostiles: altas temperaturas, ambientes corrosivos, ruido, radiaciones nucleares, alta tensión, etc.
- Generalmente, no requiere contactos.
- Bajo peso.
- Gran versatilidad geométrica.
- Compatibilidad con sistemas de comunicación de datos.
- Transmisión de datos segura.
- No emite radiofrecuencia, ni se ve afectada por ella.

Antes de entrar de lleno a hablar de los sensores de fibra óptica, repasemos algunos conceptos en relación con la propia fibra óptica.

18.1.1 FUNDAMENTOS DE LA FIBRA ÓPTICA

La luz es una onda electromagnética y como tal ocupa una zona determinada dentro del **espectro electromagnético** (figura 18.1 y anexo 3 en el CD adjunto). La región más interesante desde el punto de vista de la fibra óptica está dentro del infrarrojo cercano, concretamente entre 0,8 y 1,7 μm, ya que aquí se ven favorecidas sus propiedades de propagación.

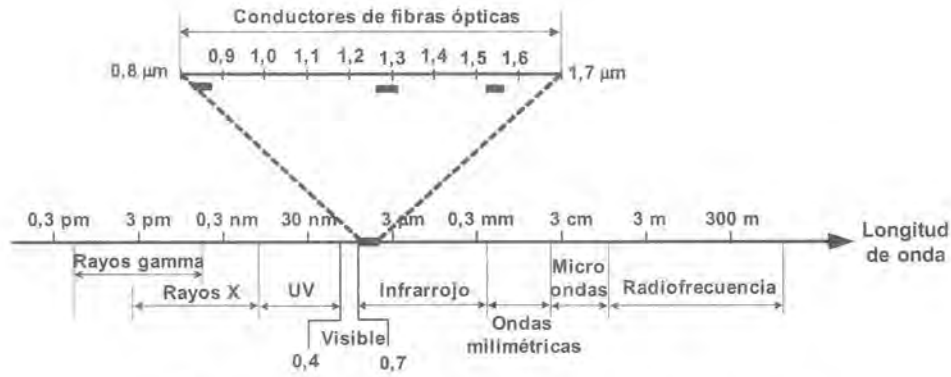


Figura 18.1. Espectro de las ondas electromagnéticas.

En general, se denomina **interferencia** a la superposición de dos o más ondas y su combinación para formar una onda única. Cuando la longitud de onda de ambas es igual y existe una diferencia de fase constante entre ellas se denominan **ondas coherentes**. Si la diferencia de fase es un múltiplo entero de λ, se produce una suma de sus amplitudes, mientras que si esa diferencia es un múltiplo entero de $\lambda/2$, una resta y, si sus amplitudes son iguales, se produce una anulación local de las ondas.

Cuando una onda luminosa pasa desde un medio isotrópico (idénticas propiedades físicas en todas las direcciones) a otro, se produce el fenómeno de la **refracción**, por el cual el rayo de luz incidente pasa al otro medio. Se define el **índice de refracción** de un medio, n, como la relación entre la velocidad de la luz en el vacío c_0 y en el medio c, es decir, $n = c_0/c$. El índice de refracción es una constante material del medio y en el vacío (≈aire) vale: $n_0 = 1$.

Si el rayo luminoso incide desde un medio 1 con índice de refracción n_1 sobre otro medio 2, con índice de refracción n_2, tal que n_1 sea mayor que n_2, una cierta parte del rayo incidente se refracta como se muestra en la figura 18.2a. La expresión matemática que describe el fenómeno de la refracción es conocida como la **Ley de la refracción de Snell**:

$$\frac{\text{sen}\theta_1}{\text{sen}\theta_2} = \frac{c_1}{c_2} \tag{18.1}$$

donde: c_1, c_2 son la velocidad de la luz en los medios 1 y 2, respectivamente. θ_1, θ_2, son los ángulos de incidencia y de refracción, respectivamente. Como $n_1 = c_0/c_1$ y $n_2 = c_0/c_2$, despejando de aquí c_1 y c_2 y sustituyendo en la expresión (18.1) se obtiene otra forma más habitual de expresar la Ley de la refracción de Snell:

$$\frac{\text{sen}\theta_1}{\text{sen}\theta_2} = \frac{n_2}{n_1} \tag{18.2}$$

Cuando el rayo luminoso incide con un ángulo igual a un valor θ_c conocido como **ángulo crítico**, el rayo refractado se propaga paralelamente a la superficie de separación de los dos medios (figura 18.2b). Como $\theta_2=90^\circ$, de la expresión (18.2) tenemos que el valor del ángulo crítico es: $\theta_c = \text{arcsen}\,(n_2/n_1)$. Para ángulos mayores que θ_c el rayo incidente es enteramente reflejado (figura 18.2c). Este fenómeno se conoce como **reflexión interna total** y es la base de la propagación de la luz en los conductores de fibra óptica. La reflexión interna total se produce sólo cuando el rayo incidente incide desde un medio con un índice de refracción mayor sobre otro de índice menor y nunca en el caso inverso.

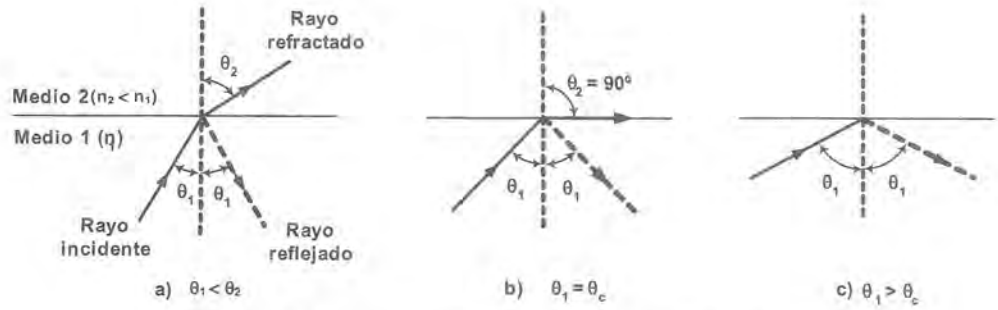


Figura 18.2. Fenómenos de reflexión y refracción de la luz.

Propagación de la luz en un conductor de fibra óptica

Un conductor de fibra óptica consiste en dos cilindros concéntricos de diámetros muy pequeños. El cilindro interior se denomina **núcleo** (*core*) y el exterior se denomina **envoltura** o **corteza** (*cladding*), siendo el índice de refracción del núcleo, n_1, ligeramente mayor que la envoltura, n_2. Estos dos cilindros suelen estar protegidos por un recubrimiento (*jacket*) como se muestra en la figura 18.3. La capacidad de guiar la luz depende de las propiedades del núcleo y de la envoltura.

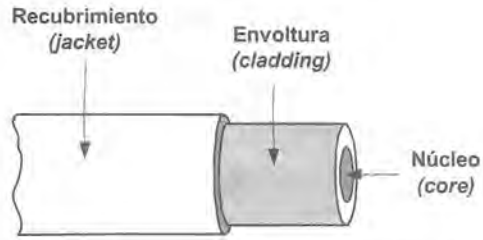


Figura 18.3. Estructura básica de una fibra óptica.

Como consecuencia de esta estructura óptica todos los rayos de luz que se reflejan totalmente en dicha superficie se transmiten guiados a lo largo de la fibra (figura 18.4).

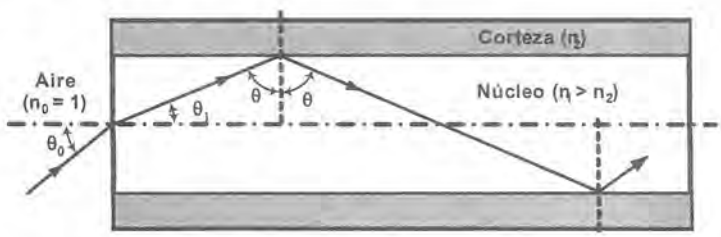


Figura 18.4. Conducción de la luz en un conductor de fibra óptica.

Para acoplar el rayo luminoso al núcleo desde el exterior de la fibra (índice de refracción n_0), el ángulo entre el rayo luminoso y el eje de la fibra se rige por la Ley de Snell:

$$n_0 \operatorname{sen}\theta_0 = n_1 \operatorname{sen}\theta_1 = n_1 \cos\theta_c = n_1 \sqrt{1 - \operatorname{sen}^2\theta_c} \qquad (18.3)$$

Sustituyendo en (18.3) la condición de ángulo crítico: $\text{sen}\theta_c = n_2/n_1$ y para un medio exterior aire, $n_0 = 1$, se tiene:

$$\text{sen}\theta_0 = \sqrt{n_1^2 - n_2^2} = AN \tag{18.4}$$

El valor $\text{sen}\theta_0$ se conoce como la **apertura numérica** (AN) y el ángulo θ_0, como **ángulo de aceptación máximo**. La apertura numérica es una medida de la capacidad de la fibra óptica para aceptar luz. La eficiencia en la captación se incrementa con el valor del parámetro AN, lo cual corresponde a mayores valores del ángulo de aceptación (figura 18.5).

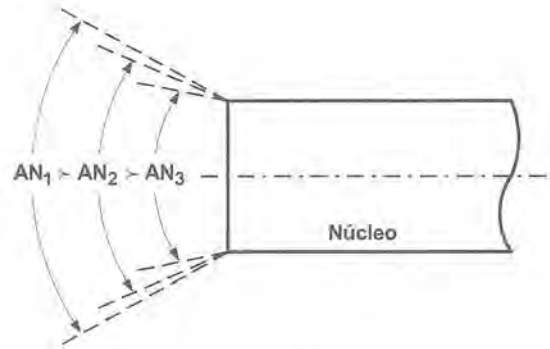


Figura 18.5. Influencia de la apertura numérica en el ángulo de aceptación.

El índice de refracción de la envoltura es ligeramente menor que el del núcleo. La **diferencia relativa de índices de refracción** Δ, se define como:

$$\Delta = \frac{n_1 - n_2}{n_1} \tag{18.5}$$

de donde: $n_2 = n_1(1-\Delta)$. Sustituyendo esta expresión en la (18.4) y simplificando resulta: $AN = n_1\left(2\Delta - \Delta^2\right)^{1/2}$.

Si $\Delta^2 << 2\Delta$, lo cual es habitual, resulta:

$$AN = n_1\left(2\Delta\right)^{1/2} = \left[2n_1\left(n_1 - n_2\right)\right]^{1/2} \tag{18.6}$$

Tipos de fibras ópticas

Atendiendo al perfil del índice de refracción del núcleo las fibras pueden ser **de salto de índice** y **de índice gradual**. Las fibras ópticas de salto de índice presentan un núcleo con un índice de refracción aproximadamente constante. En las fibras de índice gradual, el valor del índice de refracción del núcleo no es constante, sino que disminuye radialmente a medida que nos aproximamos a la envoltura. La figura 18.6 muestra el perfil de estos dos tipos de fibras.

La energía luminosa se propaga en el interior de la fibra en forma de rayos, haces o paquetes de características similares llamados **modos de propagación**. Una fibra determinada presenta un comportamiento específico respecto al conjunto de modos susceptibles de propagarse en su interior. Al objeto de caracterizar los modos de propagación de la luz en un conductor de fibra óptica se define el llamado **parámetro V**, como:

$$V = 2\pi \cdot \frac{a}{\lambda} \cdot AN \tag{18.7}$$

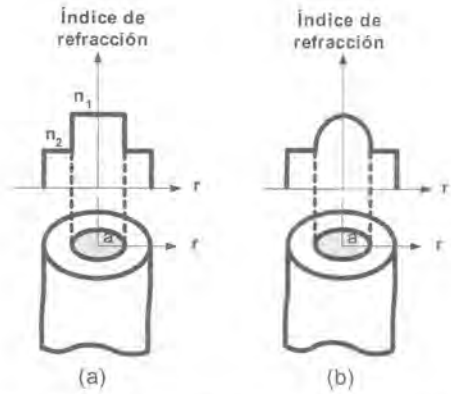


Figura 18.6. Perfil del índice de refracción: (a) fibra de salto de índice; (b) fibra de índice gradual.

donde a es el radio del núcleo de la fibra. Se tiene un valor límite de V, para el cual sólo se puede propagar en el núcleo un modo, el fundamental. Este valor límite V_c es 2,4 para los conductores de fibra óptica con perfil de salto de índice y 3,4 para los conductores de fibra óptica con perfil gradual. Así, cuando $V < V_c$ solamente se podrá propagar en el núcleo el modo fundamental, se dice que el conductor de fibra óptica es **monomodo**. Por otro lado, si $V \geq V_c$, se podrán propagar más de un modo, el conductor de fibra óptica se denomina **multimodo**. Pero, ¿cuántos modos se podrán propagar en una fibra multimodo? Si la fibra tiene un perfil de salto de índice, el número de modos será aproximadamente: $N \approx V^2/2$, mientras que en un perfil gradual es: $N \approx V^2/4$. La longitud de onda correspondiente al valor límite $V_c = 2{,}4$ se denomina **longitud de onda límite** λ_c. Si $\lambda > \lambda_c$, sólo se podrá propagar en el núcleo de la fibra un modo (el modo fundamental).

La importancia del tipo de fibra, monomodo o multimodo, es debida a un pernicioso efecto denominado **dispersión modal** que se debe a que al existir diferentes modos de propagación, un haz de luz enviado desde el extremo de una fibra llegará al otro extremo distorsionado debido a que algunos rayos recorrerán caminos más largos que otros por el interior de la fibra. La relación entre los tiempos de recorrido máximo y mínimo es directamente proporcional a la diferencia relativa de los índices de refracción Δ.

Como consecuencia de la dispersión modal, el pulso luminoso de corta duración que atraviesa el conductor de una fibra óptica de perfil de salto de índice se ensancha, reduciéndose la velocidad máxima de transmisión de datos (bits/s) o el ancho de banda de la fibra. Así, en una fibra multimodo de salto de índice, el ancho de banda apenas supera los 20 MHz; en el caso de una fibra de índice gradual, aun siendo una fibra multimodo, la dispersión modal es más pequeña que en las fibras de salto de índice, por lo que el ancho de banda llega a los gigahertzios. En el caso de un conductor de fibra óptica monomodo de salto de índice, el diámetro del núcleo es menor, así como la apertura numérica y el ángulo de aceptación, resultando difícil el acoplamiento de la luz al núcleo. En una fibra monomodo, al existir un único modo de propagación, el ancho de banda sería teóricamente infinito, pero al no ocurrir así, este se encuentra en el orden de magnitud de las decenas de gigahertzios. La figura 18.7 muestra las características de propagación de los tres tipos de fibras. En las fibras de salto de índice los rayos luminosos siguen una trayectoria quebrada cuyos vértices se encuentran en la superficie de separación del núcleo y la envoltura. En las fibras de índice gradual los rayos meridionales (rayos que viajan en un plano conteniendo el eje central del núcleo) siguen una trayectoria senoidal y los que no pasan por el eje siguen una trayectoria en forma de hélice de sección elíptica.

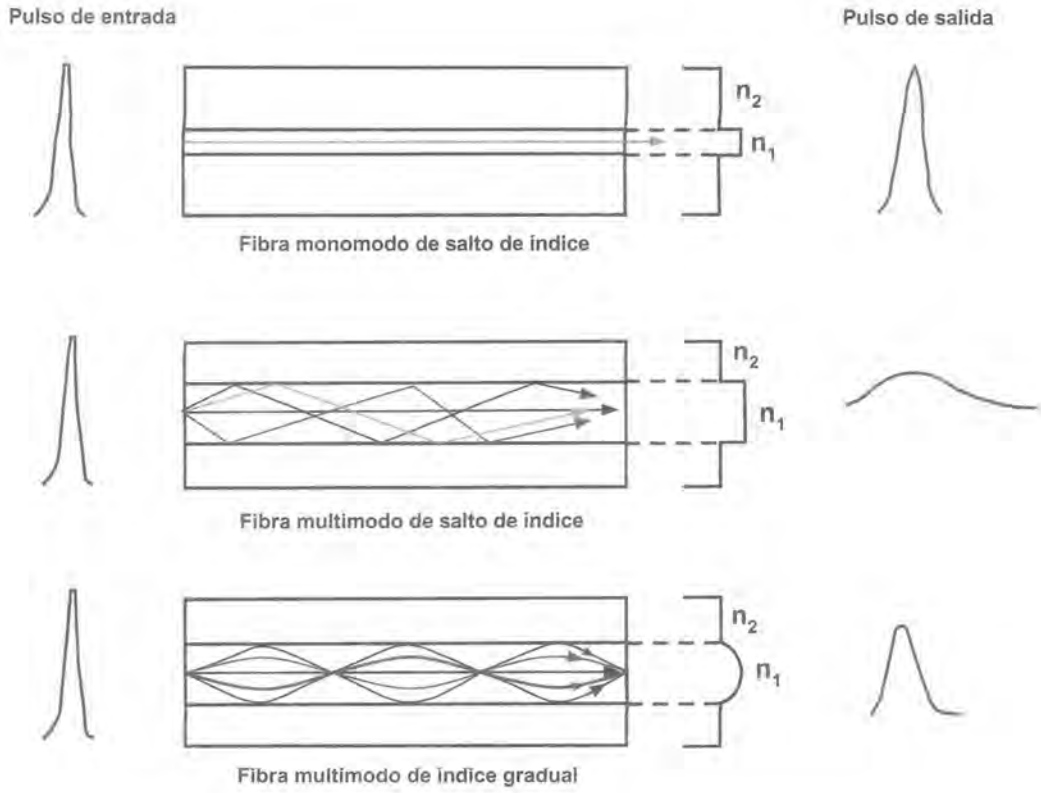


Figura 18.7. Propagación de la luz en diferentes tipos de fibra óptica.

18.1.2 CLASIFICACIÓN DE LOS SENSORES DE FIBRA ÓPTICA

La clasificación de los sensores de fibra óptica puede hacerse atendiendo a diversos criterios [18.9]:

a) Según la función realizada por la fibra óptica:
   - Sensores de fibra óptica **extrínsecos**: la fibra óptica se utiliza sólo como medio de transporte de la luz, que es modulada en un elemento externo bajo la acción de la magnitud a medir.
   - Sensores de fibra óptica **intrínsecos:** la magnitud a medir afecta a las características ópticas de la fibra. Los cambios inducidos pueden serlo en la intensidad, la fase o la polarización de la luz.

b) Según el tipo de modulación óptica empleada:
   - Sensores de fibra óptica basados en **modulación de amplitud o intensidad**. En este tipo de sensores la cantidad de luz detectada es función de las perturbaciones externas. Emplean óptica y circuitería simples. Puede emplearse una fuente de luz incoherente, tal como un LED o una fuente de luz incandescente de alta intensidad junto con fibras multimodo más baratas. Son los sensores de fibra óptica comerciales más extendidos.
   - Sensores de fibra óptica basados en **modulación de fase**. Emplean técnicas interferométricas para detectar la magnitud a medir. Presentan numerosas ventajas: flexibilidad geométrica, inmunidad a interferencias electromagnéticas, gran ancho de banda y gran resolución (por ejem-

plo, un sensor de desplazamiento basado en modulación de amplitud puede tener una resolución entre 10^{-10} y 10^{-7} m, mientras que uno basado en técnicas interferométricas puede tener una resolución de 10^{-14} m).

c) Según la aplicación:
   - Sensores de fibra óptica **físicos:** medida de temperatura, presión, desplazamiento, flujo, rotación, etc.
   - Sensores de fibra óptica **químicos**: medida de pH, análisis de gases, estudios espectroscópicos, etcétera.
   - Sensores de fibra óptica **biomédicos**: medida de glucosa, análisis de sangre, etc.

La figura 18.8 muestra un diagrama de bloques general de un sensor de fibra óptica. Generalmente, todos los sistemas tienen una fuente de luz, una interfaz con la fibra óptica, un modulador que altera la luz de forma proporcional a la perturbación (magnitud física a medir) y un detector óptico que detecta y mide los cambios en la luz de la fibra, dando una señal eléctrica en su salida que es procesada electrónicamente.

La magnitud física a medir altera los parámetros ópticos de la fibra: índice de refracción, coeficiente de absorción de la luz, dimensiones lineales. Esta alteración es debida a fenómenos físicos de diferente naturaleza: electro-óptico, magneto-óptico, piezo-eléctrico, acústico-óptico, piezo-óptico, etc.). El bloque modulador convierte el cambio en el parámetro óptico de la fibra en un cambio en un parámetro de la señal óptica que se transmite por la fibra. Se puede tener un cambio en la amplitud, la fase, la rotación del plano de polarización o en la longitud de onda de la señal óptica. Hay que tener en cuenta que un único parámetro físico puede originar diferentes tipos de medida pues puede dar lugar a distintos fenómenos físicos que modulan de diferentes formas los parámetros ópticos del medio.

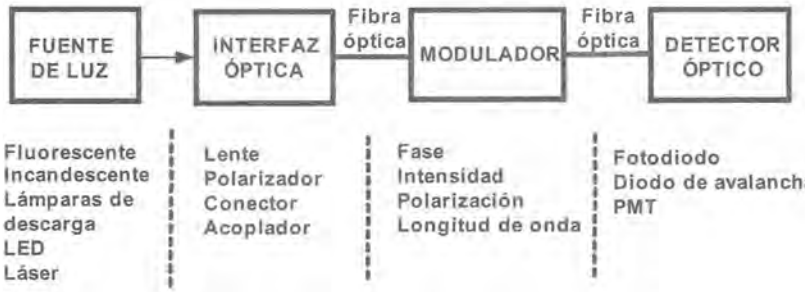


Figura 18.8. Elementos comunes a todos los sensores de fibra óptica.

18.1.3. SENSORES DE FIBRA ÓPTICA BASADOS EN MODULACIÓN DE AMPLITUD

Los sensores basados en modulación de amplitud detectan la variación de la cantidad de luz asociada con la magnitud a medir. Este tipo de sensores emplean una óptica y circuitería relativamente simple. Puede emplearse una fuente de luz incoherente, como un LED o una fuente incandescente de alta intensidad, junto con fibras multimodo. La modulación de amplitud puede llevarse a cabo empleando alguno de los siguientes mecanismos:

a) **Transmisión.** El principio de operación de estos sensores está basado en la reducción de la energía óptica transmitida por un canal entre una fuente óptica y un detector óptico. Este concepto está normalmente asociado con la medida de desplazamientos. La figura 18.9 muestra esta técnica para la medida de desplazamientos axiales y radiales, junto con las curvas de salida en función de la dis-

tancia entre las sondas de fibra óptica [18.6]. En el caso radial, la sensibilidad es mayor y para una distancia igual al diámetro de la fibra el sensor no tiene transmisión.

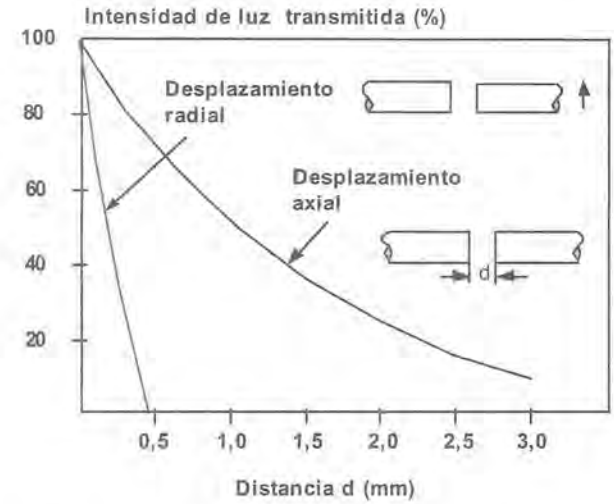


Figura 18.9. Sensor de fibra óptica de transmisión.

Se puede mejorar la sensibilidad empleando unas rejillas de absorción colocadas bien en las caras de la fibra (figura 18.10a) o entre ambas fibras (figura 18.10b). En el primer caso, se mueve una de las fibras y, en el segundo, las fibras están fijas siendo la rejilla la que experimenta un movimiento relativo entre las dos partes de la misma. Este tipo de sensor se emplea para medir la presión acústica (hidrófono).

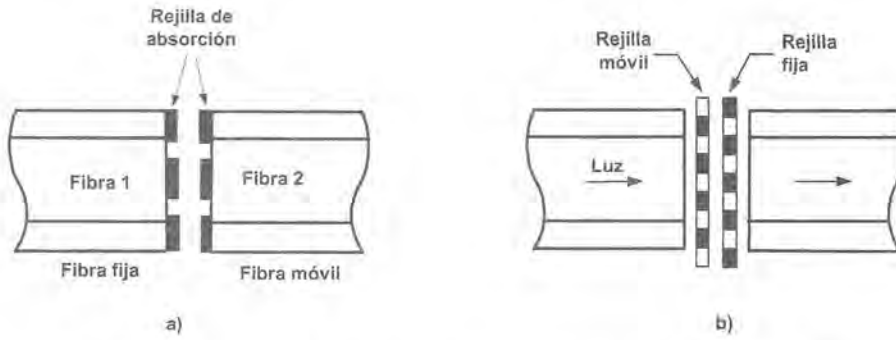


Figura 18.10. Sensor de desplazamiento radial con rejilla de absorción.

b) **Anulación de la reflexión interna total.** Una variante del concepto de transmisión es la anulación de la reflexión interna total en los límites de la propia fibra con el medio exterior. La condición de reflexión interna total de la luz está dada, como hemos visto en la figura 18.2c, por la condición: $\theta \geq \theta_c$, donde $\theta_c = \text{arc sen } (n_2/n_1)$. Basado en este concepto resulta de gran interés el sensor formado por dos fibras con una pequeña separación entre ellas como se muestra en la figura 18.11 [18.6]. Esta técnica proporciona una alta sensibilidad siendo uno de los métodos más utilizados para la detección de movimientos en estructuras mecánicas o arquitectónicas.

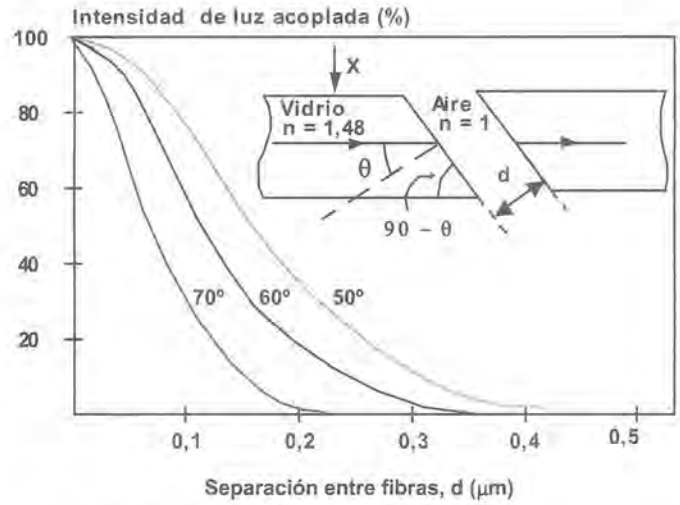


Figura 18.11. Configuración de la reflexión interna total [18.6].

c) **Reflexión.** El sensor consta de dos caminos ópticos. Una rama transmite la luz a una superficie reflectante y la otra la transmite a un detector óptico. La intensidad de la luz reflejada depende de la distancia entre la sonda y la superficie. La figura 18.12 muestra la intensidad de luz detectada en función de la distancia. La curva crece primero y desciende después según la expresión $1/d^2$, donde d es la distancia entre la punta de la fibra y la superficie de reflexión. Esta curva de respuesta y, por tanto, la exactitud, depende de la configuración interna del cable de fibra óptica. Este tipo de sensores son especialmente interesantes por su precisión, simplicidad y bajo coste. Se emplean en la medida de desplazamientos, temperatura, presión y nivel.

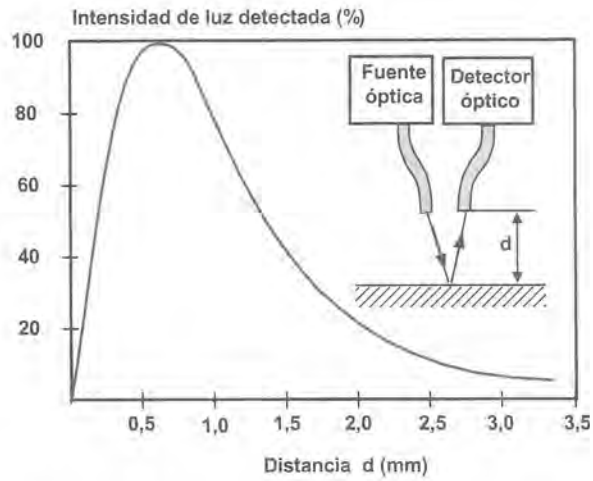


Figura 18.12. Sensor de fibra óptica de reflexión.

d) **Micro-deformaciones.** La fibra óptica sufre una pequeña deformación como consecuencia de producirse cambios en alguna magnitud física. La intensidad de luz transmitida está relacionada con el valor de dicha magnitud. La figura 18.13 muestra este tipo de sensores para el caso de medida de

desplazamiento [18.6]. La zona lineal central de la curva de respuesta es la región de medida útil. Es importante notar que el camino óptico es cerrado y, por lo tanto, son inmunes a entornos agresivos o sucios. Este tipo de sensores son muy precisos y de coste relativamente bajo. En general, la sensibilidad se incrementa con el número de puntos de deformación, y/o el espaciado entre ellos. Este mecanismo puede emplearse para la medida de presiones y de desplazamientos.

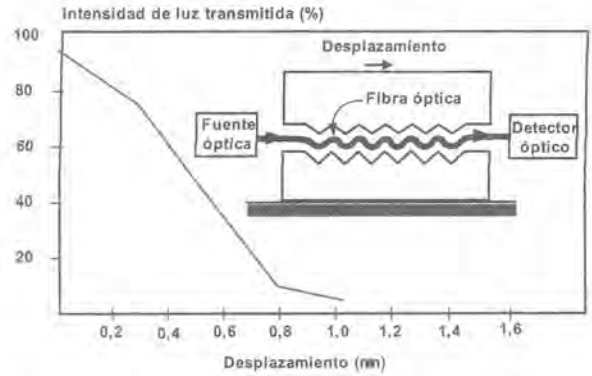


Figura 18.13. Sensor de fibra óptica basado en micro-deformaciones.

e) **Técnicas intrínsecas.** En estos sensores, es la composición química de la fibra la que da lugar a la actividad del sensor. Los principales mecanismos son: absorción, fluorescencia, cambios en el índice de refracción, polarización, etc. El fenómeno de **absorción** se tiene cuando se dopa el cristal de la fibra con tierras raras. El espectro de absorción de la fibra es sensible a la temperatura, lo cual puede ser aprovechado para medir dicha temperatura. El mecanismo de **fluorescencia** se puede obtener dopando la fibra con algunos aditivos. Se puede emplear una fuente de luz para estimular la fluorescencia o bien se puede estimular la fibra por una radiación incidente, detectándose la fluorescencia, la cual es una medida del nivel de radiación incidente (figura 18.14). Esta técnica se emplea fundamentalmente en análisis químicos: medida de la concentración de oxígeno disuelto, medida de pH (véase ejemplo de diseño en el CD adjunto), etc. En general, el empleo de la fibra óptica en los análisis químicos proporciona importantes ventajas: análisis *in situ* y en tiempo real, no afectan al proceso, el tamaño de la muestra puede ser extremadamente pequeño, la medida puede realizarse en un lugar alejado. Por el contrario se ve afectada por la luz ambiente y, en algunos casos, el tiempo de respuesta puede ser relativamente largo debido a procesos de difusión y a los tiempos de reacción.

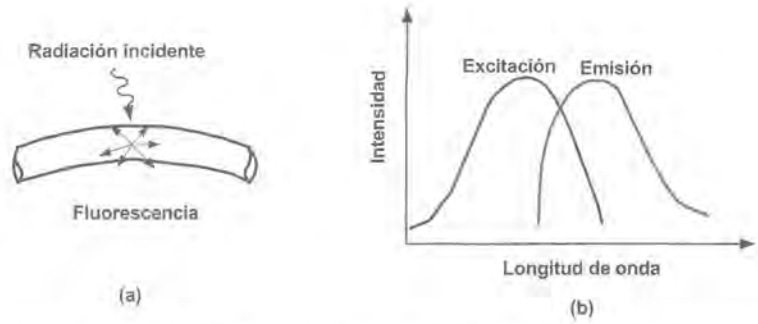


Figura 18.14. (a) Fluorescencia estimulada por radiación exterior. (b) Espectros de emisión y de excitación de una fluorescencia estimulada por una fuente de luz.

18.1.4 SENSORES DE FIBRA ÓPTICA BASADOS EN MODULACIÓN DE FASE

Estos sensores destacan por su flexibilidad geométrica, inmunidad a interferencias electromagnéticas, gran ancho de banda y, sobre todo, por su gran resolución (se han llegado a detectar desplazamientos del orden de 10^{-14} m, aproximadamente el tamaño de un núcleo atómico). Generalmente, requieren una fuente de luz coherente, dos fibras monomodo y una relativamente compleja óptica y circuitería electrónica. Las aplicaciones más extendidas son: velocidad de rotación (giróscopo), aceleración, campo magnético y presión acústica submarina (hidrófono).

La modulación de fase no puede ser detectada directamente debido a que la frecuencia de la luz es aproximadamente 10^{14} Hz y los detectores ópticos son incapaces de responder a tan altas frecuencias. Para lograr la detección de fase se emplean **técnicas interferométricas** mediante las cuales se convierte la modulación de fase a modulación de amplitud antes de la detección. Hay cuatro configuraciones de interferómetros: Mach-Zehnder, Michelson, Fabry-Perot y Sagnac [18.1].

a) **Interferómetro de Mach-Zehnder.** La figura 18.15 muestra la configuración básica del interferómetro de Mach-Zehnder. La luz procedente de un diodo láser es dividida por igual entre las fibras monomodo sensora y de referencia. El desplazamiento de fase resulta como consecuencia de cambios en la longitud y en el índice de refracción de la fibra sensora. Si las longitudes de las fibras de referencia y sensora son exactamente iguales o difieren en un número entero de longitudes de onda, los rayos recombinados en el segundo acoplador estarán en fase y la intensidad será máxima. Si los dos rayos están desfasados $\lambda/2$, el rayo recombinado tomará un valor mínimo. El segundo acoplador convierte la modulación de fase en una modulación de intensidad.

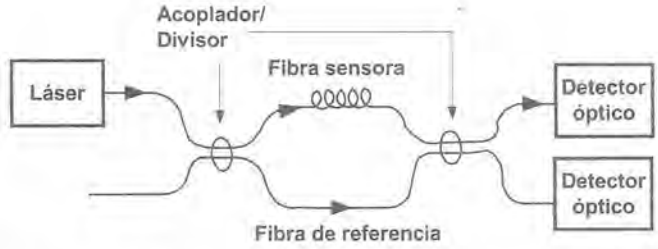


Figura 18.15. Configuración del interferómetro de Mach-Zehnder.

b) **Interferómetro de Michelson.** La figura 18.16 muestra la configuración básica del interferómetro de Michelson. Es similar al de Mach-Zehnder salvo que las fibras terminan en dos espejos, que reflejan la luz al detector a través del acoplador. Esta técnica tiene la ventaja de eliminar uno de los acopladores, sin embargo, tiene el inconveniente de que la luz se realimenta también al láser, comportándose como una fuente de ruido.

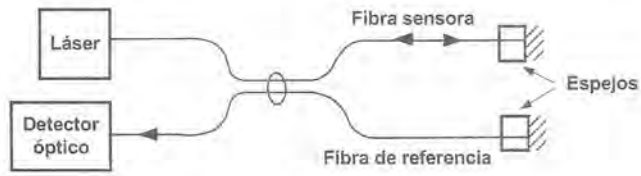


Figura 18.16. Configuración del interferómetro de Michelson.

c) **Interferómetro de Fabry-Perot.** La figura 18.17 muestra la configuración básica del interferómetro de Fabry-Perot. Este tipo de interferómetro no requiere una fibra de referencia. El rayo de luz inyectado es parcialmente reflejado al láser. La porción de luz transmitida (generalmente el 5%) llega al segundo espejo donde de nuevo una parte es reflejada y otra transmitida al detector óptico (5%). Los sensores basados en este interferómetro tienen, en general, una sensibilidad doble que las configuraciones anteriores.

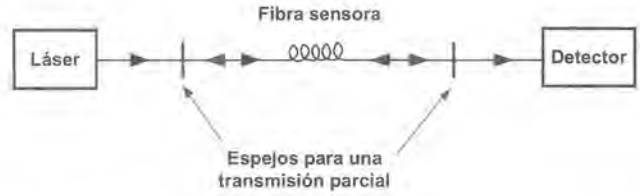


Figura 18.17. Configuración del interferómetro de Fabry-Perot.

d) **Interferómetro de Sagnac**. La figura 18.18 muestra la configuración básica del interferómetro de Sagnac. La luz del láser es inyectada, mediante el acoplador por los dos extremos de una fibra monomodo arrollada en forma de bobina. De esta forma, la luz se propaga en ambas direcciones. Mientras la bobina de fibra óptica está estacionaria no se produce un desplazamiento de fase pues la luz tarda el mismo tiempo en recorrer los dos caminos. Sin embargo, si la bobina gira en una dirección, por ejemplo en el sentido horario, el tiempo de propagación de la luz por dicho camino es menor, ya que el camino es más corto. Los dos rayos recombinados estarán por tanto desfasados. Esta técnica es empleada por los giróscopos de fibra óptica para la medida de la velocidad de rotación.

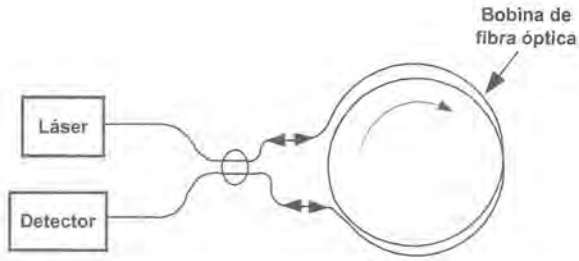


Figura 18.18. Configuración del interferómetro de Sagnac.

EJEMPLO 18.1: Medida de presión acústica mediante el interferómetro de Mach-Zehnder

La figura 18.19 muestra el esquema de un interferómetro de Mach-Zehnder usado para la medida de la presión acústica [18.2]. La luz del láser se divide entre el brazo sensor y el brazo de referencia. La fibra sensora es sometida a la presión que afecta tanto a la longitud de la fibra como al índice de refracción. Los dos rayos son recombinados detectándose la modulación de fase. Las variaciones de presión asociadas con una onda de sonido producen unas fluctuaciones de fase dadas por:

$$\Delta\phi = k\Delta(nL) = k(n\Delta L + L\Delta n) = kLn(\Delta L/L + \Delta n/n) \qquad (18.8)$$

donde k es el número de onda, n el índice de refracción del núcleo y L la longitud de la fibra.

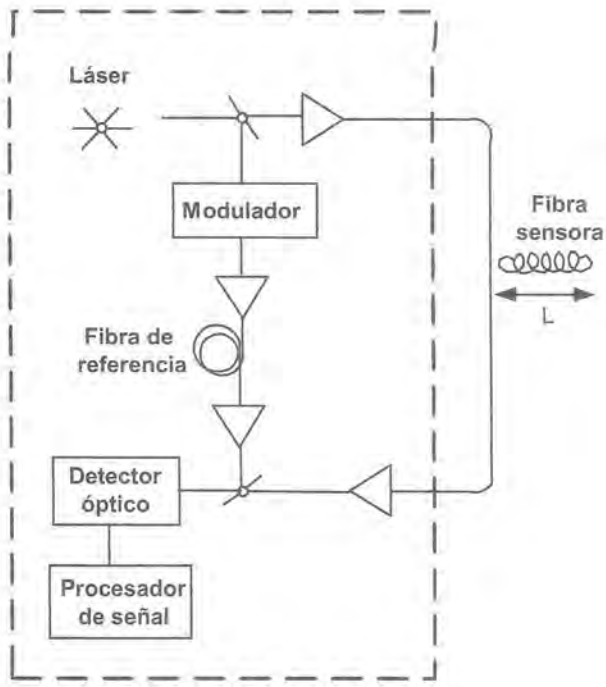


Figura 18.19

Sustituyendo en la ecuación (18.8), resulta:

$$\Delta\phi = kLn\left[\frac{1}{L}\frac{\partial L}{\partial P} + \frac{1}{n}\frac{\partial n}{\partial P}\right]\Delta P \qquad (18.9)$$

Por lo tanto:

$$\frac{\Delta\phi}{\phi\Delta P} = \left(\frac{1}{L}\right)\cdot\frac{\partial L}{\partial P} + \left(\frac{1}{n}\right)\cdot\frac{\partial n}{\partial P} \qquad (18.10)$$

EJEMPLO 18.2: Medida de la velocidad de rotación mediante el interferómetro de Sagnac

Una de las aplicaciones más interesantes de los sensores de fibra óptica basados en la modulación de fase es la medida de la velocidad de rotación (giróscopos). La mayoría de estos sensores están basados en el interferómetro de Sagnac. La figura 18.20 muestra el esquema simplificado de un sensor de fibra óptica interferométrico para la medida de la velocidad de rotación basado en el efecto Sagnac.

La luz del láser es dividida en dos por un separador y acoplada a un núcleo de fibra monomodo. La luz que emerge de los extremos de las dos fibras es combinada por el separador y detectada. En presencia de una rotación Ω, se puede demostrar (véase el problema propuesto 18.9) que:

$$\Delta L = c\,\Delta t = \frac{4AN}{c}\Omega \qquad (18.11)$$

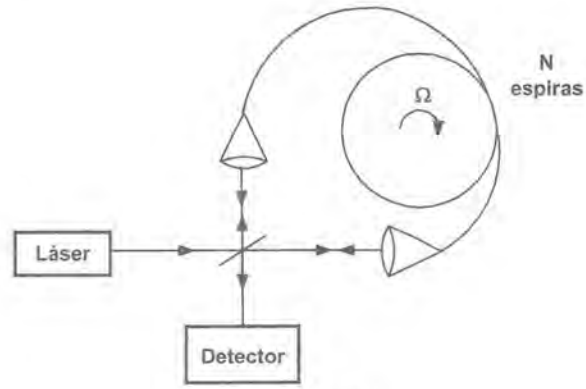


Figura 18.20

donde A es el área encerrada por el camino de luz, N el número de espiras y c la velocidad de la luz en el vacío. Despejando Δt, resulta:

$$\Delta t = \frac{4AN}{c^2}\Omega \qquad (18.12)$$

$$\Delta \phi = 2\pi \Delta t \frac{c}{\lambda_0} \qquad (18.13)$$

donde λ_0 es la longitud de onda en el vacío. Sustituyendo (18.12) en (18.13), resulta:

$$\Delta \phi = \frac{8\pi AN}{\lambda_0 c}\Omega \qquad (18.14)$$

Las principales ventajas de los giróscopos ópticos frente a los mecánicos son que no tienen partes móviles, tienen un margen dinámico mayor, bajo coste, pequeño tamaño, emplean luz. Se emplean en los sistemas de navegación de aeronaves y cohetes espaciales. El error permitido en los sensores de rotación de una aeronave puede variar entre 0,01 y 0,001°/ hora, que expresado en términos de la velocidad de rotación de la tierra (Ω_E = 15 °/h), supone entre 10^{-3} y $10^{-4}\Omega_E$.

18.2 Biosensores

La historia de los biosensores comienza en el año 1962 con el desarrollo de electrodos de enzima por el científico Leland C. Clark. Desde entonces, investigadores de diferentes campos (microelectrónica, físicos, químicos, ciencias de materiales) trabajan en el desarrollo de dispositivos biosensores cada vez más sofisticados para aplicaciones en medicina, agricultura, biotecnología, medio ambiente, etc. El crecimiento anual del mercado de los biosensores es del 60% impulsado principalmente por industrias relacionadas con el cuidado de la salud[1], el procesado de alimentos y la monitorización ambiental.

[1] En Europa Occidental se estima en un 6% el número de personas que son diabéticas y, por tanto, potenciales usuarios de biosensores de glucosa.

18.2.1 INTRODUCCIÓN

El término "biosensor" hace referencia a un dispositivo compuesto fundamentalmente por dos elementos: un elemento biológico (enzima, anticuerpo, microorganismo) y un elemento sensor (electroquímico, óptico, acústico, térmico). En esencia, en un biosensor el material biológico de reconocimiento interactúa con el analito (sustancia a detectar) y da una respuesta detectable por el elemento sensor, que convierte los cambios en las biomoléculas en una señal de salida, generalmente de tipo eléctrico. La figura 18.19 muestra la estructura general de un biosensor.

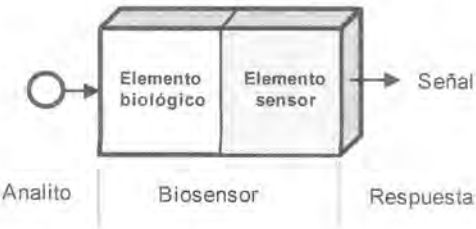


Figura 18.21. Estructura general de un biosensor.

Una de las características de los biosensores (que los distingue de otros métodos de análisis biológico) es que el analito puede ser medido de forma directa e instantánea. Otra de sus ventajas es que pueden regenerarse y volverse a usar. Además, mediante una optimización apropiada entre el ensayo biológico y el sensor primario es posible detectar concentraciones extremadamente bajas de una amplia variedad de compuestos de interés ambiental como, por ejemplo, la DBO (demanda biológica de oxígeno). Debido a su relativamente pequeño tamaño, los biosensores pueden usarse en solitario o formado parte de sistemas en donde se utilicen como detectores modulares.

Las enzimas son los materiales biológicos más utilizados. Se trata de proteínas que catalizan las reacciones bioquímicas en los seres vivos, manteniéndose sin cambios al final de dicha reacción. Básicamente destacan por su gran selectividad a un "sustrato" dado (la molécula transformada) y por favorecer la velocidad de las reacciones químicas. La velocidad máxima de la reacción es proporcional a la concentración de la enzima. La enzima ha de permanecer inmovilizada en una membrana a través de la cual el sustrato se difunde. La reacción con el sustrato da lugar a la formación de un complejo intermedio enzima-sustrato, cuya evolución y posterior ruptura da lugar a los correspondientes productos y libera la enzima (figura 18.22). Los productos de estas reacciones son detectados mediante los elementos sensores.

Figura 18.22. Principio de funcionamiento de una reacción enzimática.

Es esencial que en un biosensor se produzca un íntimo contacto entre el elemento biológico de reconocimiento y el elemento sensor, manteniendo en lo posible la estabilidad de dicho elemento biológico. La figura 18.23 muestra tres ejemplos de métodos de inmovilización. El método más sencillo es la **inmovilización por adsorción** (figura 18.23a), en el que se ponen en contacto durante un cierto tiempo las biomoléculas (B) con un material adsorbente, produciéndose la retención del biomaterial a la superficie del

sensor como resultado de fuerzas electrostáticas, enlaces de hidrógeno o de Van der Walls. En la figura 18.23b se muestra el método de **inmovilización por atrapamiento** de la biomolécula dentro de una membrana permeable sólo al analito y a los productos de la reacción de reconocimiento pero no al material biológico. La figura 18.23c representa un método químico de inmovilización basado en la formación de **uniones covalentes** entre el material biológico y la superficie del sensor. Estos métodos de inmovilización, a su vez, presentan numerosas variantes [18.10].

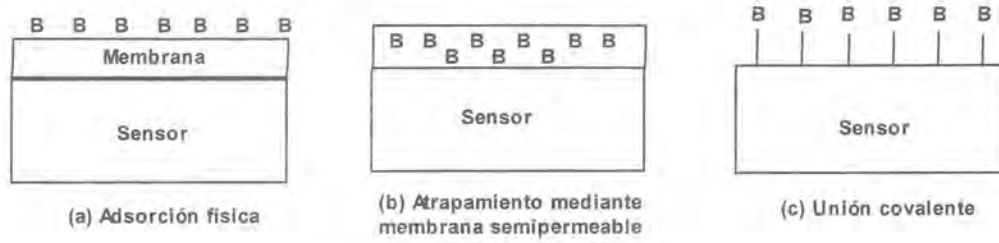


Figura 18.23. Métodos de inmovilización.

18.2.2 TIPOS DE BIOSENSORES

Atendiendo al fundamento del elemento sensor, los biosensores pueden clasificarse en: electroquímicos, ópticos, de ion selectivo, térmicos y de masa. Veamos brevemente cuál es su principio de funcionamiento.

- **Electroquímicos:** este tipo de biosensores son los más versátiles y mejor desarrollados. Utilizados principalmente en la detección de la hibridación de la secuencia del DNA, concentración de glucosa, etc., se basan en que muchas reacciones químicas producen o consumen iones o electrones, los cuales dan lugar a cambios en las propiedades eléctricas de la solución. Según cuál sea el parámetro eléctrico de medida pueden ser: potenciométricos (miden tensión), amperométricos (miden corriente) y conductimétricos (miden conductividad o resistividad). En los sensores potenciométricos se genera un potencial eléctrico en la interfase electrolito-electrodo debido a una reacción *redox* que tiene lugar en la superficie del electrodo. Los sensores conductimétricos miden los cambios en la conductividad o resistividad de un electrolito en una celda electrolítica. La tabla 18.1 compara estos tres tipos de biosensores electroquímicos.

  Los sensores electroquímicos más utilizados son los amperométricos y, de ellos, el biosensor de glucosa amperométrico es el que más éxito ha alcanzado, presentando diferentes formas y tamaños. El primer experimento que sirvió de origen a estos biosensores fue realizado por Leland C. Clark, empleando electrodos de platino para detectar oxígeno. La figura 18.24 muestra el principio de este

Características	Sensores Electroquímicos		
	Conductimétricos	Amperométricos	Potenciométricos
Parámetro medido	Conductividad	Corriente	Tensión
Tensión aplicada	Senoidal	Potencial constante	Rampa
Sensibilidad	Baja	Alta	Media
Ecuación de gobierno	Resistencia incremental	Ecuación de Cottrell	Ecuación de Nerst
Fabricación	FET + Enzima	FET + Enzima 2 electrodos	FET + Enzima Electrodo de óxido

Tabla 18.1. Tabla comparativa de los sensores electroquímicos [18.8].

experimento: la glucosa (sustrato) reacciona con la enzima glucosa oxidasa (GOD) para formar ácido glucónico, produciendo al mismo tiempo dos electrones, dos protones y reduciendo la glucosa oxidasa. Esta GOD reducida, el oxígeno circundante, los electrones y protones reaccionan para formar agua oxigenada (H_2O_2) y oxidan la GOD (la forma original). Cuanta más cantidad de glucosa se tenga, mayor es el consumo de O_2 (menor la detección) y mayor la producción de H_2O_2. La aplicación de una diferencia de potencial entre los electrodos generará a continuación una señal de respuesta.

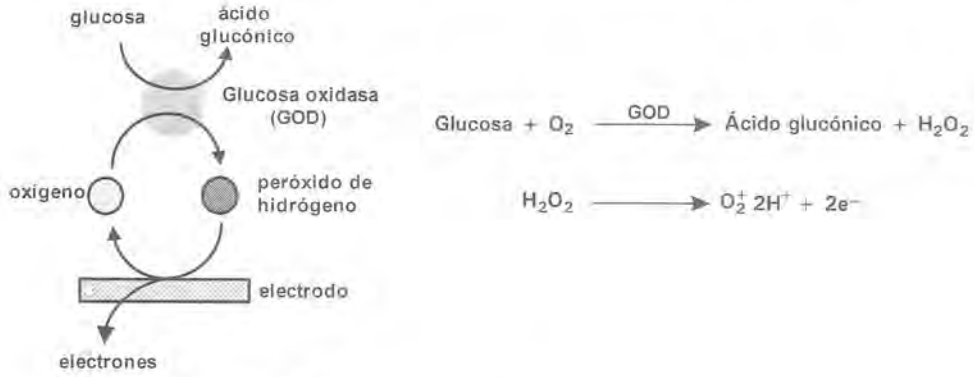


Figura 18.24. Biosensor amperométrico de glucosa.

- **Ópticos:** son la segunda familia de biosensores en importancia (después de los sensores electroquímicos), empujados por los avances en la tecnología de fibra óptica, miniaturización láser y fabricación de guías de ondas. Están basados en la interacción de radiación electromagnética con la materia, lo cual da lugar a variaciones en la intensidad, polarización o velocidad de la luz (generalmente luz visible) en el medio. Algunos biosensores son versiones miniaturizadas de espectrómetros tradicionales, mientras que otros están basados en nuevos métodos de integración óptica tales como ondas evanescentes. En el ejemplo de la figura 18.25, la luz es introducida a través de una fibra óptica recubierta con un elemento biológico. Dependiendo del número y tipo de fibras ópticas, las moléculas presentes en la superficie de la fibra absorberán ciertas longitudes de onda de la luz entrante. Se mide la luz que atraviesa (o alternativamente la que se dispersa) la capa biológica. El espectro de salida es diferente del espectro de entrada. Esta diferencia es medida y comparada con una referencia, lo cual permite determinar el nivel de analito presente.

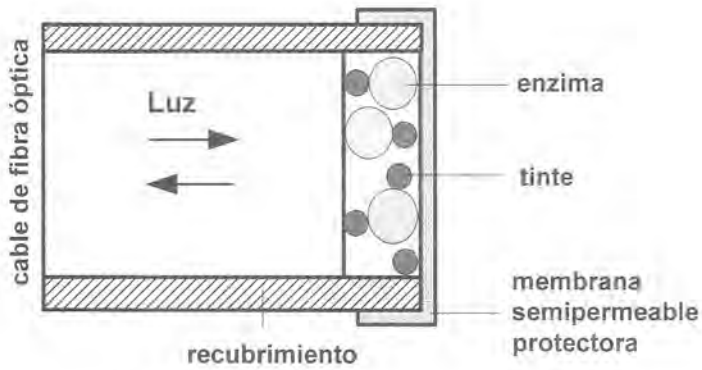


Figura 18.25. Ejemplo de biosensor óptico.

- **Ion selectivo:** son básicamente transistores de efecto de campo (FET) con una superficie sensible a ión (ISFET, *Ion-Sensible FET*). La figura 18.26a muestra el esquema de un ISFET con una capa de enzima. Es un dispositivo de tres terminales: drenaje (D), fuente (S) y un electrodo de referencia (G). El potencial eléctrico de la superficie cambia cuando los iones y la superficie interactúan. Los cambios en la densidad de carga se traducen en variaciones en la corriente de drenaje, la cual es tratada como una tensión de salida convencional V_o, como se muestra en el circuito de la figura 18.26b. Este tipo de biosensores se utilizan principalmente para la medida de pH.

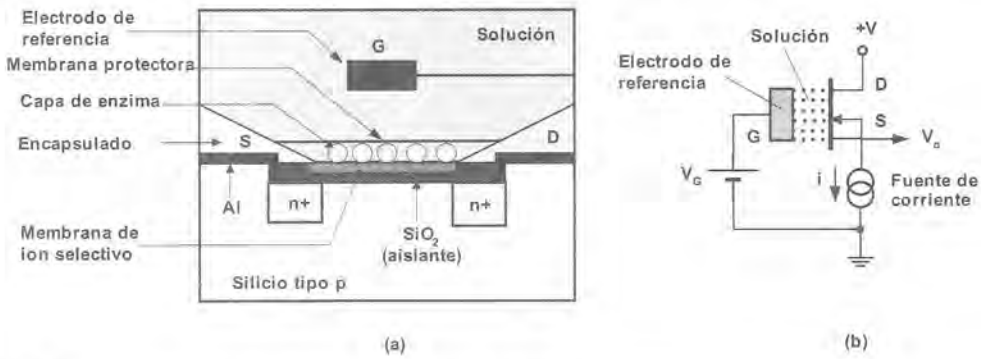


Figura 18.26. a) Transistor de efecto de campo de enzima (ENFET); b) circuito de utilización [18.4].

- **Térmicos:** este tipo de biosensores combinan enzimas con sensores de temperatura. Cuando el analito entra en contacto con la enzima se produce una reacción térmica, que es calibrada en relación con la concentración de analito. La figura 18.27 muestra un esquema simplificado de este tipo de biosensores. Contiene una capa aislante para reducir las pérdidas de calor. El termistor está rodeado por una enzima inmovilizada en una matriz. Este sensor de temperatura junto con otro de referencia se conectan a los brazos de un puente de Wheatstone.

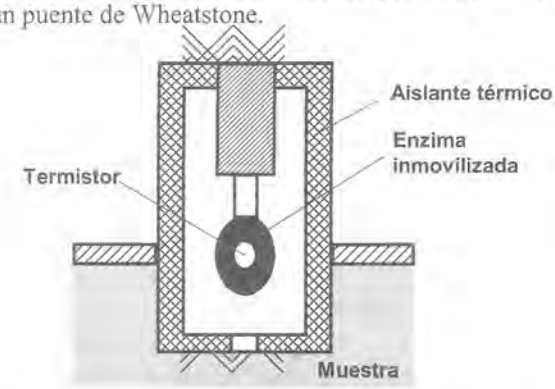


Figura 18.27. Esquema de un biosensor térmico de enzima inmovilizada [18.2].

- **De masa:** están formados por un sensor de onda acústica acoplado a un anticuerpo[2]. Cuando las moléculas del analito (antígeno) se unen a la membrana, la masa de esta cambia, originando un cambio en la frecuencia de resonancia del sensor dado por la expresión:

[2] Los anticuerpos son proteínas solubles producidas por los organismos vivos como reacción a la presencia de sustancias extrañas, antígenos, que son marcadas para su posterior eliminación por dicho organismo. La gran especificidad de las uniones antígeno-anticuerpo es lo que lleva a la utilización de estos sistemas para el desarrollo de biosensores [18.10].

$$\Delta f = f_o \Delta_m S_m \qquad (18.15)$$

donde: f_o es la frecuencia natural de oscilación, Δ_m la masa añadida por unidad de área y S_m la sensibilidad, la cual depende del diseño, material y frecuencia de trabajo (longitud de onda) del sensor acústico. Son biosensores extremadamente sensibles y con un margen dinámico bastante grande.

Tratamiento electrónico de la señal

La señal eléctrica proporcionada por el biosensor es generalmente de muy bajo nivel y está superpuesta sobre una señal de relativamente alto nivel y ruidosa. El acondicionamiento normalmente requiere restarle una señal base de referencia procedente de un sensor sin membrana biocatalizadora, amplificar la señal diferencia resultante y realizar un filtrado para eliminar el ruido procedente de interferencias electrónicas o el generado por los propios componentes electrónicos. La naturaleza relativamente lenta de la respuesta del biosensor facilita el problema del filtrado del ruido. A continuación, la señal es cuantificada por un convertidor analógico-digital y llevada a un dispositivo procesador donde los datos son convertidos a unidades de concentración, llevados a un display y/o almacenados. La figura 18.28 muestra un diagrama de bloques genérico para el tratamiento de la señal.

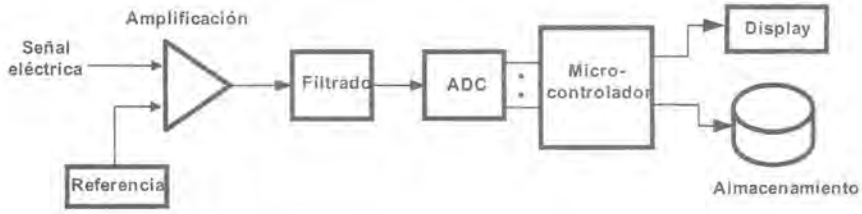


Figura 18.28. Tratamiento electrónico de la señal de salida de un biosensor.

Campos de aplicación

Los biosensores encuentran aplicación en diferentes sectores: clínico, agroalimentario, protección, medio ambiente y defensa. Dentro del campo clínico el electrodo de glucosa oxidasa es el más empleado debido a la necesidad de medir la glucosa en sangre y en orina para el diagnóstico de la diabetes. Como ejemplo, la figura 18.29 muestra un biosensor amperométrico de glucosa desechable desarrollado por P. Yu y S. Dong

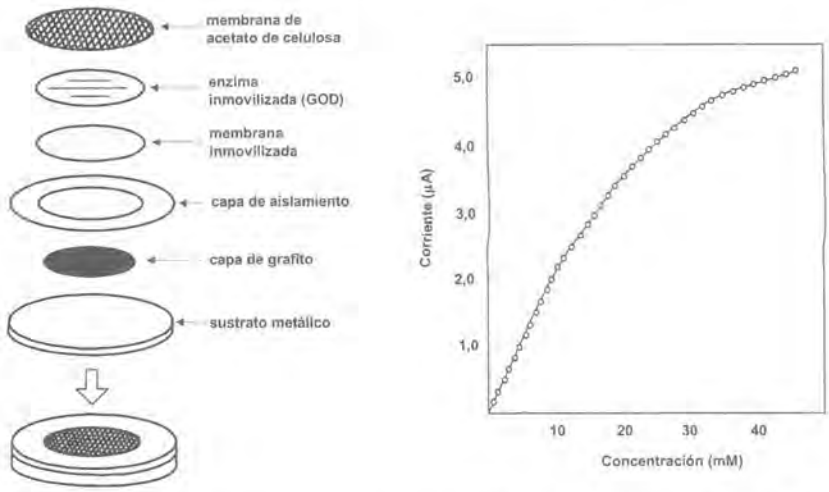


Figura 18.29. Detalles constructivos y curva de calibración de un biosensor de glucosa desechable.

[18.13]. Tiene unas dimensiones de 8 mm de diámetro y 0,88 mm de espesor. La capa de aislamiento está situada en los electrodos de grafito lo cual permite filtrar la interferencia de otras sustancias (ácido ascórbico, ácido úrico) permitiendo sólo el paso de H_2O_2 y O_2. La capa exterior de acetato de celulosa proporciona una barrera para la interferencia con otras sustancias. Se muestra también la curva de calibración del biosensor. Se observa cómo la relación es bastante lineal hasta una concentración de glucosa de aproximadamente 25 mMol/l.

Resumen

- Si bien en un principio las aplicaciones de la fibra óptica se orientaron hacia las comunicaciones, pronto se vio la posibilidad de llevarlas al campo de los sensores, empujados por los continuos desarrollos en la tecnología de la fibra óptica.
- Los sensores de fibra óptica destacan por tener, en general, una mayor sensibilidad que los sensores tradicionales, buen comportamiento en entornos hostiles (altas temperaturas, ambientes corrosivos, ruido, radiaciones nucleares, alta tensión), generalmente no requieren contactos, bajo peso, gran versatilidad geométrica, compatibilidad con sistemas de comunicación de datos, transmisión de datos segura, no emiten radiofrecuencia, ni se ven afectados por ella.
- A pesar de las ventajas anteriores su elevado coste hace que, de momento, su campo de aplicación se vea reducido a aplicaciones donde no prima el factor económico.
- Los sensores de fibra óptica pueden clasificarse atendiendo a diversos criterios, si bien en todos ellos lo que en último término se produce es una modulación de alguno de los parámetros ópticos: amplitud, fase o polarización.
- Los sensores de fibra óptica basados en modulación de amplitud son los más sencillos (no precisan radiación coherente ni demandan fuentes de luz ni detectores ópticos especiales) y los que más implantación comercial tienen en la actualidad.
- Los sensores de fase o interferométricos destacan por su extraordinaria sensibilidad y su gran flexibilidad geométrica. La detección de la señal acústica submarina (hidrófonos) y la medida de la velocidad de rotación (giróscopos) son dos buenos ejemplos de este tipo de sensores.
- Los biosensores son un tipo especial de sensor químico en los que en esencia, un material biológico de reconocimiento (enzima, anticuerpo) interactúa con el analito (sustancia a detectar) y dan una respuesta detectable por el elemento sensor, que convierte los cambios en las biomoléculas en una señal de salida, generalmente de tipo eléctrico. Se caracterizan por su gran sensibilidad y selectividad.
- Los biosensores más utilizados son, sin duda, los electroquímicos, debido principalmente a que la señal de salida es eléctrica, los límites de detección son suficientemente amplios, siendo la instrumentación necesaria relativamente simple y con bajo coste.

Problemas propuestos

18.1 Razone si al aumentar el diámetro de una fibra óptica se incrementa la cantidad de luz que puede capturar.

18.2 ¿Por qué es más interesante para la fibra óptica la región comprendida entre los 800 y 1700 nm?

18.3 Represente sobre una fibra óptica la propagación de los rayos en una fibra monomodo y otra multimodo de salto de índice. Compare las ventajas e inconvenientes de estos tipos de fibras de cara a su utilización como un canal óptico.

18.4 Represente una sección de fibra óptica de salto de índice y dibuje en ella la propagación de dos rayos de luz, uno de los cuales incide desde el aire con un ángulo menor que el ángulo crítico y otro con un ángulo mayor.

18.5 Un conductor de fibra óptica con perfil gradual tiene las siguientes características: diámetro del núcleo 100 μm; apertura numérica 0,2; longitud de onda 1 μm. Determine si se trata de un conductor de fibra óptica monomodo o multimodo.

18.6 Una fibra multimodo de salto de índice tiene una diferencia relativa de índices de refracción del 1% y un índice de refracción del núcleo de 1,46 y el número de modos de propagación a la longitud de onda de 1,2 μm es de 1000. Determine el diámetro del núcleo de la fibra.

18.7 Determine la longitud de onda límite de una fibra monomodo de salto de índice si el índice de refracción de núcleo es de 1,46, el radio de 5 μm y la diferencia relativa de índices es 0,5%.

18.8 La tabla siguiente muestra las dimensiones típicas de dos conductores de fibra óptica con perfil de salto de índice. Determine para cada uno: (a) ángulo crítico; (b) ángulo de aceptación; (c) diferencia relativa de índices de refracción; (d) parámetro V si $\lambda = 0{,}85$ μm; (e) número de modos. ¿En cuál de ellos resulta más difícil el acoplamiento de luz?

Dimensiones típicas	Fibra multimodo	Fibra monomodo
Diámetro del núcleo	100 μm	5 μm
Diámetro de la envoltura	140 μm	100 μm
Índice de refracción del núcleo	1,48	1,46
Índice de refracción de la envoltura	1,46	1,45

18.9 La figura 18.30 representa una espira de fibra óptica de radio R que gira a la velocidad angular Ω con respecto a un eje perpendicular al plano del disco. Deduzca la expresión 18.11 del efecto Sagnac en el vacío, en el que están basados los giróscopos ópticos. Suponga que del punto 1 parten en sentidos opuestos dos fotones, de forma que para $\Omega = 0$, ambos tardan el mismo tiempo en recorrer la espira y volver al punto de partida. En cambio para $\Omega \neq 0$ el tiempo que emplean es distinto.

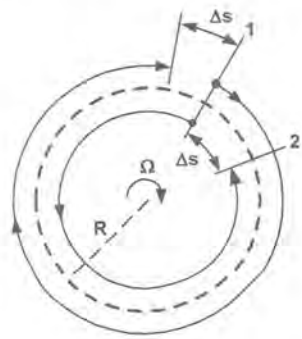


Figura 18.30

19

Criterios para la selección de sensores

19.1 Medida de la temperatura
19.2 Medida de las principales variables mecánicas
19.3 Medida de magnitudes eléctricas
19.4 Medida de magnitudes ópticas

19.1 Medida de la temperatura

19.1.1 INTRODUCCIÓN

Existen infinidad de procesos industriales en los que resulta imprescindible controlar la temperatura, e incluso en el ámbito doméstico es una de las magnitudes de mayor interés.

La elección del sensor adecuado para cada aplicación no es evidente, por lo que seguidamente se compararán los más habituales:

- Resistencias metálicas (RTD).
- Termopares.
- Termistores (NTC).
- Sensores de silicio.
- Sensores de radiación.

En general, la mayor dificultad consiste en decidir entre RTDs y termopares, o entre NTCs y sensores de silicio, por lo que se comparan entre sí por separado. Los sensores de radiación tienen aplicaciones más definidas, por eso no suelen entrar en competencia con otros sensores de temperatura. De todas formas, se tratarán brevemente los diferentes métodos y sensores para la medida de radiación en aplicaciones termométricas.

19.1.2 RTD vs TERMOPAR

Industrialmente los dos sensores de temperatura más utilizados son los termopares y las RTDs. Aunque existen RTDs de diferentes metales (cobre, níquel, etc.), las más empleadas son las de platino debido a su linealidad, estabilidad y repetibilidad, por lo que son las únicas que se considerarán en el presente apartado. Dentro de las RTDs de platino, existen dos tipos fundamentales: las de película y las bobinadas. Las bobinadas tienen mejores características pero su precio es más elevado. Las de película presentaban en sus orígenes problemas de estabilidad; sin embargo, en la actualidad, estos problemas han sido prácticamente eliminados y se han convertido en una buena alternativa a las bobinadas.

En cuanto a los termopares, existen varios tipos en función de los materiales que los constituyen cuyas diferencias más importantes son el campo de medida para los que resultan adecuados y el ambiente al que van destinados (oxidante, reductor, etc.).

A continuación, se comparan las RTDs y los termopares atendiendo a aspectos tales como el campo de medida, la estabilidad, etc.

Campo de medida

El campo de medida constituye uno de los factores más importantes para decantarse a favor de un termopar. Las RTDs de platino son adecuadas para medir entre –200°C y 850°C como máximo, aunque, en la práctica, el límite superior es de unos 600°C. Los termopares permiten alcanzar, en términos generales, temperaturas mayores (hasta 1800°C, dependiendo del tipo). Industrialmente, es muy habitual tener que medir temperaturas fuera del alcance de las RTDs por lo que resulta obligada la elección del termopar.

Intercambiabilidad

Tanto las RTDs como los termopares están perfectamente caracterizados. En el caso de las primeras, las normas muestran la resistencia frente a la temperatura y en el caso de los termopares muestran la tensión que proporcionan en función de la temperatura cuando la unión de referencia está a 0°C. Sin embargo, los componentes comerciales presentan tolerancias; la intercambiabilidad hace referencia a la desviación en la medida de la temperatura debido a la tolerancia del sensor. La norma EN60751 [19.2] define dos clases de tolerancia para las RTDs de platino (tabla 19.1): A y B. La clase A tiene una tolerancia menor que la B.

Clase de tolerancia	Tolerancia (°C)
A	$0{,}15 + 0{,}002 \cdot \lvert t \rvert$
B	$0{,}3 + 0{,}005 \cdot \lvert t \rvert$

Tabla 19.1. Clases de tolerancia de las RTDs de platino ($|t|$ Temperatura en valor absoluto).

Por lo tanto, cuando se reemplaza una RTD por otra de la misma clase, la medición de la temperatura puede sufrir variaciones (tabla 19.2) si no se procede a una recalibración.

En el caso de los termopares, la norma UNE-EN 60584-2 [19.3] define tres clases de tolerancia (1, 2 y 3). Las tolerancias de intercambiabilidad de los diferentes termopares en función de la temperatura se muestran en la tabla 19.3. Comparándolas con las tolerancias de las RTDs se puede observar que cuando la temperatura es elevada las tolerancias pueden ser similares. No obstante, las tablas están definidas para termopares nuevos, pero a largo plazo los termopares muestran desviaciones mayores que las RTDs.

Temperatura	Tolerancia de temperatura	
	Clase A	Clase B
–200°C	±0,55°C	±1,3°C
–100°C	±0,35°C	±0,8°C
0°C	±0,15°C	±0,3°C
100°C	±0,35°C	±0,8°C
200°C	±0,55°C	±1,3°C
300°C	±0,75°C	±1,8°C
400°C	±0,95°C	±2,3°C
500°C	±1,15°C	±2,8°C
600°C	±1,35°C	±3,3°C
700°C	-	±3,8°C
800°C	-	±4,3°C

Tabla 19.2. Tolerancia de intercambiabilidad de las Pt100 a diferentes temperaturas (para la clase A no se aplica la expresión de la tabla 19.1 para temperaturas superiores a 650°C). (Elaborada según la norma UNE-EN 60751.)

T(°C)	Tolerancias (±°C)														
	J			K y N			T			R y S			B		
	Clase			Clase			Clase			Clase			Clase		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
–200	-	-	-	-	-	-	-	-	3,0	-		-	-	-	
–100	-	-	-	-	-	-	-	-	1,5	-		-	-	-	
0	1,5	2,5	-	1,5	2,5	-	0,5	1,0	1,0	1,0	1,5	-	-	-	
100	1,5	2,5	-	1,5	2,5	-	0,5	1,0	-	1,0	1,5	-	-	-	
200	1,5	2,5	-	1,5	2,5	-	0,8	1,5	-	1,0	1,5	-	-	-	
300	1,5	2,5	-	1,5	2,5	-	1,2	2,3	-	1,0	1,5	-	-	-	
400	1,6	3,0	-	1,6	3,0	-	-	-	-	1,0	1,5	-	-	-	
500	2,0	3,8	-	2,0	3,8	-	-	-	-	1,0	1,5	-	-	-	
600	2,4	4,5	-	2,4	4,5	-	-	-	-	1,0	1,5	-	-	1,5	4,0
700	2,8	5,3	-	2,8	5,3	-	-	-	-	1,0	1,8	-	-	1,8	4,0
800	-	-	-	3,2	6,0	-	-	-	-	1,0	2,0	-	-	2,0	4,0
900	-	-	-	3,6	6,8	-	-	-	-	1,0	2,3	-	-	2,3	4,5
1000	-	-	-	4,0	7,5	-	-	-	-	1,0	2,5	-	-	2,5	5,0
1100	-	-	-	-	-	-	-	-	-	1,0	2,8	-	-	2,8	5,5
1200	-	-	-	-	-	-	-	-	-	1,3	3,0	-	-	3,0	6,0
1300	-	-	-	-	-	-	-	-	-	1,6	3,3	-	-	3,3	6,5
1400	-	-	-	-	-	-	-	-	-	1,9	3,5	-	-	3,5	7,0
1500	-	-	-	-	-	-	-	-	-	2,2	3,8	-	-	3,8	7,5
1600	-	-	-	-	-	-	-	-	-	2,5	4,0	-	-	4,0	8,0
1700	-	-	-	-	-	-	-	-	-	-	-	-	-	4,3	8,5

Tabla 19.3. Tolerancias de intercambiabilidad de los termopares a diferentes temperaturas (los campos no marcados no están incluidos en la norma UNE-EN 60584-2).

Estabilidad

Una de las características más destacables de las RTDs es la excelente estabilidad y constituye una de las principales razones para escogerlas en una aplicación determinada. Si un proceso debe realizarse a una temperatura específica, podrá determinarse en un laboratorio la resistencia de la RTD a esa temperatura y apenas variará a lo largo del tiempo, lo que permitirá garantizar la estabilidad del proceso.

Un termopar es menos estable cuando se expone a temperaturas medias o altas. En aplicaciones críticas los termopares deben retirarse y comprobarse periódicamente bajo condiciones controladas.

Robustez

Uno de los puntos considerados tradicionalmente débiles de las RTDs es que son un tanto delicadas para los ambientes industriales. Las bobinadas están hechas de un fino hilo de platino que se puede romper con relativa facilidad. El problema en las de película reside en el contacto de los hilos de conexión con el platino. De todas formas, se ha progresado mucho en este sentido y hoy en día las RTDs tienen cada vez mayor resistencia mecánica.

Los termopares son muy robustos y presentan una duración mayor que las RTDs. Resultan muy adecuadas en ambientes sujetos a elevadas vibraciones y a riesgo de choques.

Tamaño

Industrialmente, tanto termopares como RTDs se encuentran normalmente protegidos por una vaina metálica (normalmente de acero inoxidable) denominándose al conjunto sonda termométrica (figura 19.1). De todas formas, el termopar es un sensor de menor tamaño que la RTD, lo que permite utilizar vainas de diámetro mucho más reducido. Por ejemplo, existen comercialmente agujas hipodérmicas que incluyen un termopar en su interior. También cabe la posibilidad de no utilizar vainas protectoras permitiendo alcanzar lugares de muy difícil acceso cuando se emplean termopares, aunque la estabilidad disminuye. En general, los termopares resultan más adecuados que las RTDs cuando se desea realizar una medida de la temperatura en un punto muy concreto, debido a su menor tamaño.

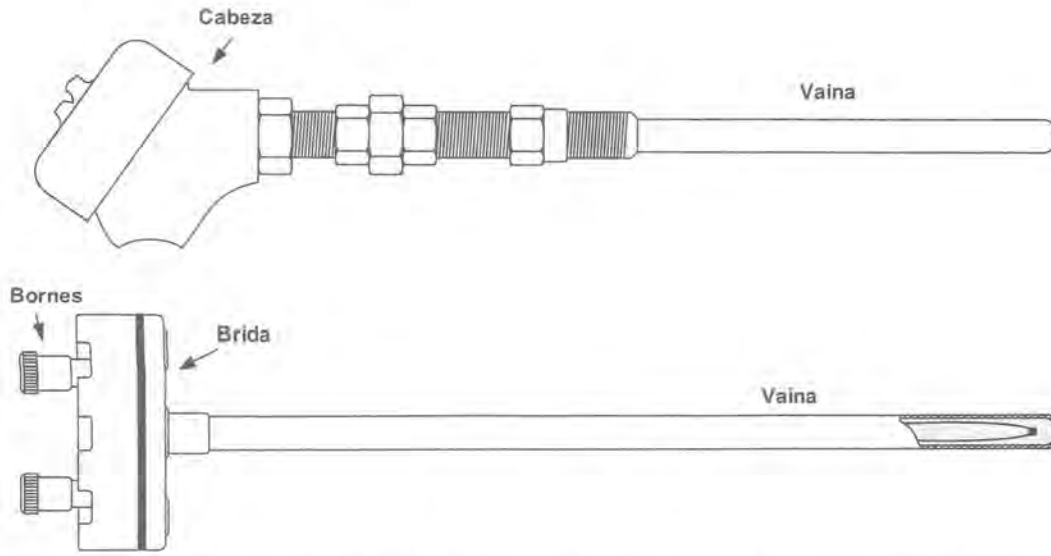


Figura 19.1. Montajes típicos de las sondas termométricas para termopares y RTDs.

Otros aspectos

Para una misma configuración una RTD puede costar entre 4 y 10 veces más que un termopar de un metal base. Por otra parte, el tiempo de respuesta de un termopar es bastante más reducido; al menos unas tres veces más rápido.

La elección entre un termopar y una RTD para una determinada aplicación no es sencilla. La práctica demuestra que el primero es el sensor más utilizado en ambientes industriales (en Estados Unidos aún más que en Europa [19.19]). De todas formas, las RTDs de película presentan ventajas crecientes y son, cada vez, una mejor alternativa a los termopares en aplicaciones tradicionalmente reservadas a estos últimos.

19.1.3 NTC vs SENSORES DE SILICIO

La elección entre una NTC y un sensor de silicio resulta cada vez más difícil. En principio, las NTCs presentan un campo de medida útil más amplio. Existen NTCs que permiten medir temperaturas de hasta 500°C aunque la mayor parte no sobrepasan los 100-150°C. Además, las mayores ventajas se obtienen cuando el campo de medida es reducido debido a que pueden realizarse circuitos de acondicionamiento muy sencillos con una elevada linealidad en la tensión de salida (pese a la marcada no linealidad del sensor) y a que la estabilidad es muy elevada en estos casos. Por ello, también resultan muy adecuadas para efectuar la compensación de temperatura de otros sensores.

Entre los puntos más débiles de la NTC se encuentra la mala intercambiabilidad que suele obligar algún tipo de calibración del sistema cada vez que se sustituyen. Si bien es cierto que existen comercialmente NTCs con tolerancias reducidas (hasta del 1% o incluso menores), su coste llega a ser incluso superior a las RTDs por lo que resultan poco competitivas.

Los sensores de silicio son una excelente alternativa a las NTCs aunque su coste es generalmente mayor. Presentan la ventaja de proporcionar directamente la salida en tensión y con una sensibilidad bastante elevada por lo que cada vez resulta más habitual su utilización en lugar de las NTCs.

Finalmente, en la tabla 19.4 se muestra una comparación entre RTDs, termopares, NTCs y sensores de silicio.

Características	Sensor				
	RTD de platino de película	RTD de platino bobinada	Termopar	Termistor	Silicio
Coste del sensor	Moderado a bajo	Moderado	Bajo	Bajo a moderado	Bajo
Coste del sistema	Moderado	Moderado	Alto	Bajo a moderado	Bajo
Campo de medida	−200 a 750°C (560°C máx., típicamente)	−200 a 850°C (600°C máx., típicamente)	−270 a 1800°C	−100 a 500°C (125°C máx., típicamente)	-40 a 125°C
Intercambiabilidad	±0,1%, ±0,3°C	±0,06%, ±0,2°C	±0,5%, ±2°C	±10%, ±0,2°C	±1%, ±3°C
Estabilidad	Excelente	Excelente	Pobre	Moderada	Moderada
Sensibilidad	0,39%/°C	0,39%/°C	40 μV/°C	−4%/°C	10 mV/°C
Sensibilidad relativa	Moderada	Moderada	Baja	Muy elevada	Moderada
Linealidad	Excelente	Excelente	Moderada	No es lineal	Moderada
Pendiente	Positiva	Positiva	Positiva	Negativa	Positiva
Susceptibilidad al ruido	Baja	Baja	Alta	Baja	Baja

Tabla 19.4. Comparación entre diferentes sensores de temperatura (elaborada según [19.12]).

19.1.4 SENSORES DE RADIACIÓN

Los termómetros de radiación miden la temperatura a partir de la energía que emiten todos los materiales en función de su temperatura en la región infrarroja, fundamentalmente.

Se utilizan cuando no se pueden emplear los termopares, RTDs u otros por alguno de los siguientes motivos:

- Cuando es necesario efectuar la medida sin contacto. Por ejemplo, porque resulta difícil el acceso al punto de medida debido a que el objeto está en movimiento y no resulta posible mantener una conexión entre el sensor y el sistema de control o de visualización de datos, o porque no resulta conveniente realizar contacto alguno como puede suceder en la industria alimentaria o en determinadas aplicaciones médicas.
- Cuando el objeto está rodeado de un fuerte campo electromagnético que provoca interferencias. Por ejemplo, en un horno de inducción.
- Cuando se requiere un tiempo de respuesta muy reducido. Por ejemplo, algunos termómetros clínicos de aparición en el mercado relativamente reciente se basan en la medida de la radiación gracias a lo que permiten medir la temperatura en menos de un segundo.
- Cuando la temperatura es muy elevada y queda fuera del alcance de los termopares.

Básicamente, están constituidos por un sistema óptico para realizar el enfoque del objetivo en el sensor de radiación, el propio sensor y el sistema electrónico de acondicionamiento de señal y visualización. En muchos casos, incluyen un LED o un diodo láser que ilumina la zona de medición para permitir que el usuario conozca fácilmente la zona que está midiendo.

Uno de los mayores inconvenientes de los termómetros de radiación es que la medición depende de la emisividad del cuerpo cuya temperatura se pretende determinar. Básicamente, se conocen tres tipos de materiales en función de la emisividad:

- Cuerpo negro: el que presenta una emisividad igual a uno. Una cavidad negra con un pequeño orificio por el cual la radiación se comporta, aproximadamente, como un cuerpo negro. De todas formas, se trata de un concepto puramente teórico; en la práctica, todos los materiales presentan emisividades menores que uno.
- Cuerpo gris: el que presenta una emisividad constante pero menor que uno. La mayor parte de los materiales están incluidos dentro de esta categoría. Las emisividades de muchos materiales están comprendidas entre 0,85 y 0,95. Sin embargo, algunos materiales como el oro, la plata y el aluminio presentan emisividades muy pequeñas (entre 0,02 y 0,04) lo que hace muy difícil la medida de su temperatura por radiación.
- Cuerpo no gris: el que presenta una emisividad variable en función de la longitud de onda.

Un termómetro de radiación viene ajustado para una emisividad determinada (normalmente 0,95), por lo que si se mide la temperatura de un cuerpo cuya emisividad es diferente, la lectura debe corregirse. Para ello se puede recurrir a tablas donde se muestran las emisividades [19.23] o bien a la determinación experimental empleando otro termómetro de referencia.

Otros inconvenientes de importancia son los siguientes:

- Son mucho más costosos que otros termómetros como los basados en RTDs o termopares y presentan una menor exactitud. Además, no se dispone de tablas de calibración como las recogidas en las normas para estos dos últimos sensores.

- La medición puede presentar desviaciones de importancia en presencia de vapor de agua, polvo o determinados gases entre el sensor y el objeto a medir que absorban radiación dentro de la zona de medición.
- Las tablas de emisividad de los diferentes materiales están obtenidas realizando la medida de forma perpendicular al objeto. Para otros ángulos es necesario realizar una corrección experimental si resulta necesario mantener una determinada exactitud.

Por todas estas cuestiones, la utilización correcta de los termómetros de radiación suele requerir una cierta experiencia. También debe prestarse especial cuidado a la medición de materiales aparentemente transparentes (véase ejemplo 19.1).

EJEMPLO 19.1

El vidrio es un cuerpo no gris por lo que su emisividad depende de la longitud de onda. A longitudes de onda por debajo de 3 μm es muy reducida (también depende del grosor); por ello, es transparente en la zona visible del espectro. Sin embargo, la emisividad es muy elevada a longitudes de onda grandes, pudiendo llegar a 0,95-0,97 alrededor de 8-10 μm.

Si se pretende medir la temperatura del filamento de una bombilla incandescente mediante un termómetro de radiación, sería necesario que este midiera únicamente la radiación emitida a longitudes de onda donde el vidrio sea transparente. Por ejemplo, sería adecuado un termómetro que trabajase en el entorno de 2 μm [19.23].

Si lo que se pretende medir es la temperatura de la ampolla de vidrio, sería necesario un termómetro que midiera la radiación a longitudes de onda para las que el vidrio fuese opaco. Por ejemplo, en el entorno de 10 μm.

En cuanto a los sensores para la medida de la radiación hay que tener en cuenta que la longitud de onda a la que tiene lugar el pico de emisión de radiación de un cuerpo depende de la temperatura y es tanto más pequeña cuanto mayor es esta última. Por ejemplo, a 0°C la longitud de onda de pico es de unos 10 μm, mientras que a 1000°C es de 2,3 μm. Por lo tanto, la respuesta espectral del sensor es uno de los parámetros más importantes para realizar su selección. Algunos de los sensores más comunes son los siguientes (figura 19.2):

- Fotorresistencias: algunas de las más utilizadas son las de PbS y las de InSb. Estas últimas presentan una respuesta espectral más amplia y centrada en longitudes de onda más largas por lo que resultan adecuadas para medir temperaturas más reducidas.
- Fotodiodos: los de Si y Ge responden únicamente a longitudes relativamente cortas por lo que son adecuados para medir temperaturas elevadas; generalmente mayores de 600-1000°C. Los de InAs permiten medir temperaturas más reducidas.
- Sensores piroeléctricos: además de presentar una respuesta espectral muy amplia, lo que permite campos de medida grandes, la sensibilidad es relativamente grande. Uno de los mayores inconvenientes reside en que únicamente son apropiados para la medida de señales alternas por lo que se necesita trocear la radiación incidente mediante un elemento mecánico (*chopper*).
- Termopilas: con las termopilas se mide el incremento de la temperatura de una superficie negra muy absorbente de radiación (idealmente, un cuerpo negro), debido a la incidencia de la radiación procedente del objeto cuya temperatura se pretende medir. Las termopilas están constituidas por decenas de termopares en serie para aumentar la sensibilidad del conjunto puesto que el incremento de tempera-

tura del objeto es, generalmente, reducido, y de forma que la unión caliente de todos ellos se pone en contacto con la superficie negra. Por supuesto, se debe proceder a la compensación de la unión fría como en cualquier termopar.

Si bien las termopilas son conocidas desde hace mucho tiempo, su utilización ha crecido de forma espectacular en los últimos años debido a la aparición de las termopilas de película delgada que están fabricadas con técnicas similares a los semiconductores. Presentan una respuesta espectral muy amplia lo que permite campos de medida muy grande, su tamaño es reducido, la sensibilidad es aceptablemente buena y el coste es relativamente bajo.

- Bolómetros: son termómetros de resistencia (RTDs o NTCs) configurados para medir radiación de un modo similar a las termopilas.

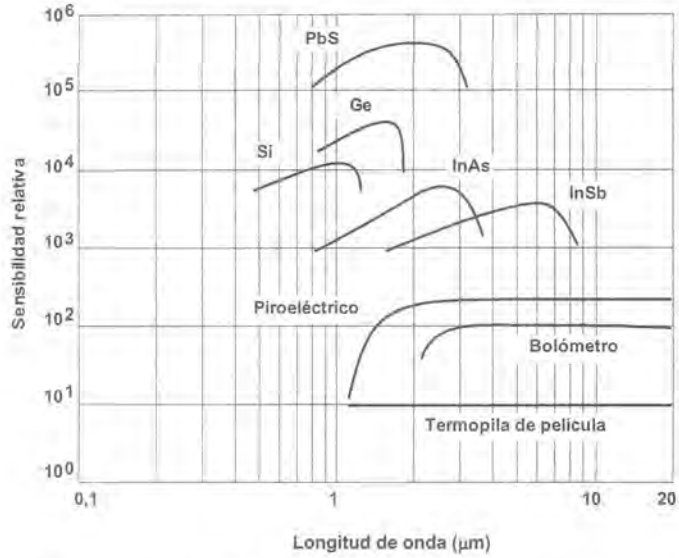


Figura 19.2. Respuesta espectral de diferentes sensores utilizados para termómetros de radiación (tomada de [19.23]).

En función de la porción del espectro que se emplea para obtener la medida de la temperatura, los sistemas se clasifican en:

- Pirómetros de radiación total.
- Pirómetros de radiación parcial.
- Pirómetros ratiométricos.

En los primeros se mide con un sensor de respuesta espectral muy amplia, desde 300 nm hasta 2,5 μm-20 μm lo que permite recoger una porción importante de la radiación térmica emitida por el objeto. Los sensores empleados son normalmente de tipo piroeléctrico o termopilas, y los campos de medida más habituales suelen ser 0-1000°C y 500-900°C. La exactitud ronda el 1%.

En los pirómetros de radiación parcial la medición se efectúa en una zona estrecha, normalmente menor de 1 μm, bien porque el propio sensor tiene una respuesta espectral reducida (como puede ser el caso de un fotodiodo de silicio) o porque se limita mediante filtros ópticos. El inconveniente es que mucha de la energía que llega al sensor no se transforma en una señal eléctrica por lo que la sensiblidad disminuye y obliga a un diseño mucho más cuidadoso y costoso que en el caso de los pirómetros de radiación total. Pero presentan

notables ventajas puesto que, si se trabaja a longitudes de onda a las que no absorban radiación los posibles gases presentes entre el instrumento y el objeto a medir, se evitan las interferencias. Además, se puede realizar un diseño más adaptado a la aplicación mejorando la exactitud. Algunos campos de medida de instrumentos comerciales son –40/600°C, 0/1000°C, 600/3000°C y 500/2000°C presentando exactitudes que van desde el 0,25% en márgenes reducidos hasta el 2% sobre el fondo escala.

En los pirómetros ratiométricos se mide la energía recibida en dos bandas estrechas obteniéndose el valor de la temperatura a partir del cociente entre ambos resultados. También se denominan pirómetros de dos colores puesto que si se efectuase la medición en la zona visible cada banda correspondería a un color diferente. Resultan muy adecuados para la medición de la temperatura de materiales no grises, en los que la emisividad varía en función de la longitud de onda. Además, son menos sensibles a las variaciones de la emisividad de un mismo material en función del acabado superficial y a las interferencias debidas a la presencia de gases absorbentes. Algunos campos de medida típicos de sistemas comerciales son 900/3000°C y 50/3700°C y la exactitud típica es del 1-2% sobre el fondo de escala.

19.2 Medida de las principales variables mecánicas

En muchos sistemas industriales suele ser necesaria la medida de múltiples variables mecánicas, y esta consideración se puede hacer extensiva a otros sistemas como los domésticos y, así, en la vida diaria nos vamos encontrando con muchas variables de tipo mecánico que se están midiendo con sistemas electrónicos que nos rodean y que suelen incluir sensores para hacer esas medidas.

De todas las variables que caben dentro del epígrafe "mecánicas" —que resulta muy amplio— no pretendemos hacer un recuento de cada una de ellas y asignar los sensores más apropiados para cada caso ya que eso desbordaría el alcance que se pretende de este libro convirtiéndolo en una interminable lista de variables y sensores. Por el contrario, se tratará de hacer hincapié en las variables que sean más importantes o más complejas con la intención de que el lector vaya extrayendo las consideraciones necesarias para llegar a ser capaz de elaborar criterios de selección propios para cualquier otro caso no contemplado aquí.

Si en los capítulos anteriores hemos hecho una clasificación orientada al sensor, ahora vamos a clasificar en función de las variables y podemos hablar de:

a) Variables que se caracterizan por depender del lugar que un ente ocupa en el espacio y/o de las variaciones de ese lugar. Así tenemos:
   1. Las que sólo dependen del lugar tales como la posición y el nivel como variables analógicas y la proximidad y presencia como variables digitales.
   2. Las que dependen de la variación de lugar que serán las respectivas derivadas de la variable que determina la posición y que serán: la velocidad (lineal y angular) tanto de sólidos como de líquidos, la aceleración (lineal y angular) y la vibración.

b) Variables que dependen de la fuerza tales como la propia fuerza, la presión, el peso y el par.

c) Variables mixtas tales como el flujo másico o el caudal que combina variables de velocidad y de masa.

No cabe duda que quedan más en el tintero y que algunas de ellas pueden tener importancia para casos concretos pero aquí vamos a restringirnos a unas pocas por resultar extraordinariamente habituales en muchos sistemas de instrumentación. En concreto, vamos a tratar las siguientes:

- Presencia y proximidad.
- Posición.

- Velocidad.
- Aceleración y vibración.
- Fuerza y peso.
- Presión.
- Caudal.
- Nivel.

A estas dedicaremos los siguientes apartados e indicaremos cuáles son los criterios a emplear para seleccionar cualquier sensor aplicable a la medida de cada una de esas variables.

19.2.1 MEDIDA DE PRESENCIA Y PROXIMIDAD

Aunque en esencia ambas medidas sean casi lo mismo ya que se trata de indicar con una variable digital si hay algún "objeto" en el campo de alcance del sensor, el concepto de presencia hace referencia a un campo de acción relativamente grande, mientras que la proximidad se refiere a un alcance mucho más restringido aunque puede ser muy difícil saber cuándo se trata de una variable de proximidad y cuándo es de presencia (si hay dudas, en el fondo, es que tampoco importa mucho qué nombre le damos) por lo que hablaremos de sensores que sean capaces de detectar la existencia de un objeto dentro de su radio de acción.

Cuando se trata de detectar la presencia/ausencia de un objeto en un área amplia, la medida se podrá hacer mediante los cambios que el objeto introduzca en ese área para lo cual tenemos dos tipos de sistemas de medida:

a) **Sensores pasivos** que monitorizan la situación de determinada variable y detectan los cambios producidos por el objeto sobre esa variable. Es un tipo muy común de medida empleado en sistemas de seguridad de bajo coste o en sistemas de iluminación activados mediante la presencia de una persona en el área que deba ser iluminada. La variable más habitual de medida es la luz en la zona del espectro infrarrojo y las zonas que pueden cubrir este tipo de sensores dependen de los sistemas ópticos —habitualmente lentes de Fresnel[1]— que se empleen. La luz recogida por el sistema óptico se lleva a un detector infrarrojo (un fotodiodo) que la recibe y mide los cambios que se produzcan. Estos sistemas suelen incorporar filtros de luz que evitan interferencias.

   También son pasivos los sensores de imagen que recogen una imagen más o menos nítida y que la procesan para detectar la presencia de cualquier objeto. Para este tipo de sensores se emplean CCDs o matrices con unos pocos fotodiodos; en el primer caso se precisa un procesado de imagen para detectar el objeto o su movimiento, mientras que en el segundo caso el procesamiento es más simple y corresponde a la detección de cambios de señal en cualquiera de los fotodiodos.

   Finalmente, es posible pensar en sensores de sonido o ultrasonidos (micrófonos) basados en el efecto piezoeléctrico capaces de detectar los ruidos que hipotéticamente ocasionaría un objeto desplazándose por el área en la que el micrófono tiene alcance.

b) **Sensores activos** que producen una señal y recogen su eco. El eco de la señal dependerá de la presencia o ausencia de objetos en el campo de acción y/o de la velocidad con que se desplace. Para llevar a cabo la detección del objeto se puede utilizar la emisión de sonido (ultrasonido en realidad) y técnicas de impulso-eco o emplear un emisor de luz que envíe un haz del tipo que sea y comprobar si se recibe el eco o si llega a un receptor. En el primer caso, se tienen detectores de presencia de ultrasonidos que suelen permitir cubrir áreas relativamente grandes, mientras que los otros o

[1] La lente de Fresnel fue una realización de Augustin Fresnel que llevó a la práctica en 1820 una idea del conde de Bufón que permitía reducir el coste y el peso de una lente mediante el uso de un conjunto de anillos concéntricos.

bien cubren áreas más pequeñas o se trata de detectores de barrera y sólo permiten saber cuándo un objeto corta un haz de luz entre una fuente emisora y un receptor, habitualmente un fotodiodo.

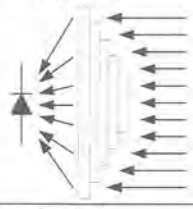
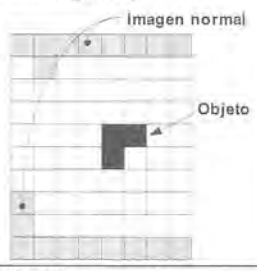
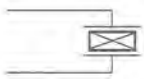
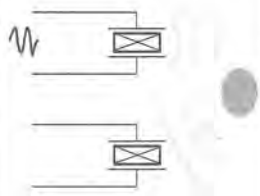
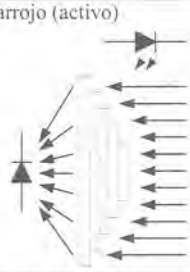
Tipo de sensor	Modo de trabajo	Ventajas e inconvenientes	Aplicaciones
Infrarrojo (pasivo) 	Detección de cambios en los niveles de luz que recibe un **fotodiodo** y que provienen de un área bastante grande.	+ Bajo coste + Uso sencillo + Alcance de algunos metros – Poco preciso – Puede ser interferido – No detecta movimientos lentos – Precisa algo de luz	– Seguridad básica – Control de iluminación
Cámara (pasivo) 	Procesamiento digital de la imagen que capta una cámara y que normalmente usa algún tipo de **CCD**.	+ Bastante preciso + Puede permitir más funciones: Identificación Velocidad Monitorización + Alcance modificable – Alto coste – Diseño complejo – Precisa iluminación	– Seguridad – Control de accesos – Monitorización
Micrófono (pasivo) 	Detección del nivel de ruido que produce un objeto al moverse mediante un dispositivo **piezoeléctrico**.	+ Sencillo – Muy poco preciso – Fácil de interferir – Poco sensible	– Seguridad básica
Ultrasonidos (activo) 	Mediante técnicas de impulso-eco se detecta la presencia de objetos. Emplea un emisor y un receptor de tipo **piezoeléctrico**.	+ Bastante sensible + Permite conocer distancia + No precisa movimiento + Sencillo – Le afectan las vibraciones – Bastante omnidireccional – Alcance de pocos metros	– Seguridad – Control de acceso – Control de iluminación
Infrarrojo (activo) 	Similar al infrarrojo pasivo pero ahora se incluye un sistema de iluminación infrarroja en el propio sistema.	– Bajo coste + Uso sencillo + Alcance de algunos metros – Poco preciso – Puede ser interferido – No detecta movimientos lentos	– Seguridad básica – Control de iluminación
Barrera infrarroja (activo)	Se establece una línea de luz entre un emisor (LED) y un receptor (**fotodiodo**) que permite saber cuándo un objeto la interrumpe.	+ Bajo coste + Uso sencillo + Alcance variable desde dos o tres metros hasta decenas. + Muy preciso – Requiere un entorno limpio	– Seguridad básica – Control de accesos y de puertas – Uso industrial

Tabla 19.5. Algunos sensores de presencia y sus características.

En la tabla 19.5 se recogen algunas de las alternativas para realizar la detección de presencia entendiendo como tal la detección de un objeto relativamente alejado del sistema que lo detecta en una zona también amplia sin contacto físico con él.

La detección de proximidad presenta características similares a la detección de presencia pero con la salvedad de que la distancia es más pequeña y el uso de los sistemas de detección de proximidad suele quedar restringido a aplicaciones industriales de control de movimiento, detección de objetos en cadenas de producción, recuento, etc. La primera posibilidad de detectar proximidad es mediante el uso del contacto de tal forma que el objeto actúa sobre una palanca que cierra o abre un contacto, tal y como se comportan los finales de carrera. En la figura 19.3 se muestra un esquema muy simple de este tipo de dispositivo que es muy popular y cuyo uso está extraordinariamente extendido por su sencillez.

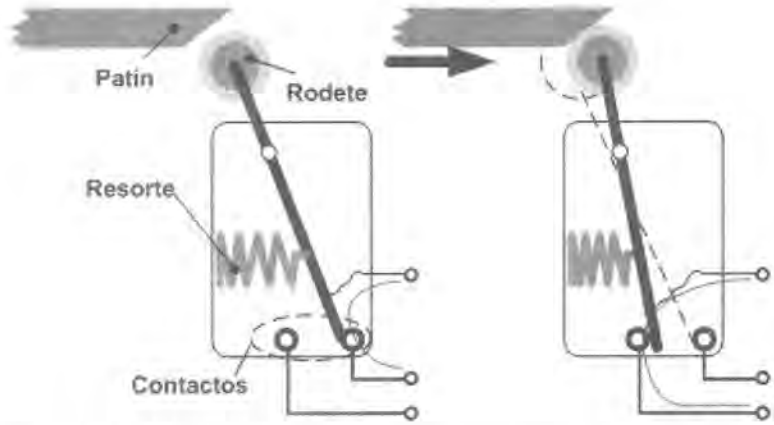


Figura 19.3. Final de carrera, uno de los sensores más populares y sencillos.

No obstante, la robustez mecánica de un sistema como el final de carrera no es su principal virtud y la vida útil de este tipo de sistemas se mide en número de contactos y aunque su diseño suele ser muy bueno, la necesidad de contacto físico entre el objeto y el sensor lo desaconseja en bastantes ocasiones y lo invalida en otras.

Las posibilidades de detectar la proximidad a un objeto sin necesidad de contacto físico dependen de la naturaleza del objeto y la influencia que este pueda tener sobre alguna variable más sencilla de medir; así, aunque puedan emplearse técnicas diversas —y seguro que a poco que pensemos se nos pueden ocurrir aún más— lo cierto es que, a nivel real, se tienen las siguientes formas de detección de proximidad de objetos:

a) **Mediante cambios en el campo magnético.** Los sensores inductivos generan un campo magnético y cualquier objeto metálico (puede ser de un material ferromagnético o no, aunque en el caso de que lo sea, el efecto es mayor aún) lo modifica, lo que se traduce en un cambio en algún valor de corriente o de tensión en el circuito del dispositivo inductivo.

b) **Mediante cambios en el campo eléctrico.** Los sensores capacitivos son capaces de producir un campo eléctrico que será modificado por cualquier objeto en las proximidades del sensor: si el objeto es metálico, el cambio será evidente, mientras que si es no metálico, la magnitud del cambio dependerá de la diferencia entre la constante dieléctrica del material y la del aire.

c) **Mediante ruptura de un haz de luz.** Como en los sensores de barrera que se pueden aplicar a estos casos pero en distancias más cortas; la barrera suele estar constituida por un conjunto emisor-receptor y un espejo adecuado situado a una distancia no muy grande y que devuelve el haz al conjunto emisor-receptor, denominándose como detectores de reflexión en espejo.

d) **Mediante reflejo de luz.** Estos sensores, denominados normalmente como detectores de reflexión en objeto, son similares a los infrarrojos activos que se usaban para detección de presencia pero ahora su alcance es mucho menor y la iluminación resulta muy direccional.

e) **Mediante ecos de ultrasonidos.** Igual que lo que se utilizaba en los detectores de presencia pero también, en este caso, con un alcance mucho menor.

En la tabla 19.6 se recogen algunos de los principales detectores de proximidad con sus características más destacadas.

Sensor	Alcance[2]	Limitaciones	Usos
Final de carrera	0 (por contacto)	– Exige contacto – Duración limitada	– Automatismos simples – Seguridad – Uso industrial (decayendo)
Inductivo	0,1 a 20 cm	– Sólo objetos metálicos – Objetos grandes (el tamaño influye en el alcance) – Selectivo	– Automatismos – Recuento de piezas – Seguridad – Uso industrial
Capacitivo	0,1 a 3 cm	– Objetos grandes (el tamaño influye en el alcance) – Corto alcance – Selectivo	
Barrera de luz	0,1 a 10 m	– Sólo detectan en una línea – Entornos limpios – Poco selectivo	
Reflexión en espejo	0,1 a 3 m	– Sólo detectan en una línea – Entornos limpios – Poco selectivo	
Reflexión en objeto	0,05 a 1 m	– Poco selectivo – El alcance depende del objeto a detectar – Entorno limpio	
Ultrasonidos	0,1 a 3 m	– Poco selectivo – El alcance depende del tamaño del objeto – Sensibles a vibraciones – Poco direccional	

Tabla 19.6. Detectores de proximidad.

Cuando se elige uno de estos detectores, el criterio de selección más habitual es el *alcance de detección* pero se ha de tener en cuenta que no siempre interesa un alcance grande ni ser capaz de detectar objetos muy pequeños; es más, en muchas ocasiones interesa que el alcance sea bajo (sólo se detecta lo que está muy próximo y el resto de objetos no interfiere) o sólo los objetos que son significativos. Para comprenderlo, supongamos el simple ejemplo de un detector de proximidad inductivo en una planta industrial que deba detectar la presencia de una parte voluminosa de una máquina que se desplaza: sería absurdo que detectase

[2] El alcance de cualquiera de estos sensores es muy variado y dependerá de circunstancias diversas entre las que la más importante es la naturaleza del objeto a detectar. Por ello, los datos facilitados sólo deben ser tomados como una referencia general y nunca como una limitación exacta; para conocer las prestaciones reales deberá acudirse a los datos de los fabricantes cuyas páginas Web se enlazan desde el CD adjunto.

cualquier tornillo o pequeño trozo de metal que pudiera haber cerca. Por tanto, estos sistemas deben ser selectivos, es decir, dar señales válidas lo que implica que no indiquen que hay objeto cuando no lo hay ni viceversa. De esta forma se pone de manifiesto la importancia de la selectividad de un detector.

Como otros criterios de selección están los referidos a la naturaleza del objeto a detectar y sus características que pueden empujar hacia una decisión u otra según las limitaciones establecidas en la tabla 19.6.

19.2.2 MEDIDA DE LA POSICIÓN

La medida de la posición de determinados objetos es un caso bastante habitual en el mundo de la instrumentación industrial; se presentan varios casos generales, algunos de los cuales trataremos en este apartado:

a) Medida de la **posición angular** que es muy típica en todos los sistemas que incluyen servomotores y tiene una especial aplicación en robótica, aunque no se restringe a ella ya que muchos accionamientos precisan conocer el ángulo de giro de una determinada pieza (posición angular).

b) Medida de la **posición lineal** que es algo menos normal que la anterior pero que no por ello está poco presente: el concepto general de medida de la distancia a un objeto es muy amplio y cae fuera del alcance del texto por lo que nos vamos a restringir a la medida en distancias relativamente cortas. En particular, haremos especial hincapié en los casos de medida de distancia en accionamientos lineales que es un caso muy típico a nivel de control industrial.

A continuación, veremos cómo efectuar la medida de las posiciones angulares y de la distancia a objetos cercanos.

Posición angular

La posición angular se puede medir de muchas formas pero nos vamos a restringir a los modos más típicos de medida que pasan por el empleo de potenciómetros o de *encoders*; aunque estos últimos pueden ser de diversos tipos (ópticos, magnéticos, capacitivos) los más habituales son los de tipo óptico. Otras opciones para la medida de posición angular son las que pasan por el empleo de *resolvers* o de dispositivos inductivos y capacitivos aunque su aplicación actual sea muy reducida.

La medida de la posición angular es una tarea muy normal en muchos procesos industriales ya que existe multitud de casos en que el accionamiento es un motor que debe situarse en determinada posición, constituyendo lo que se denomina un servomotor de posición. Todos estos sistemas incorporan algún sistema de medida de posición angular a la salida del eje (normalmente tras el reductor). Con la masiva implantación de robots y brazos robotizados[3] en las cadenas de montaje el empleo de servomotores de posición se ha hecho mucho más abundante. En la figura 19.4 se muestra el diagrama de bloques básico de un servomotor de posición.

Para la medida de posición hay un gran número de sensores aplicables y, aunque en la mayoría de los casos la solución es el *encoder* óptico, no faltan aplicaciones en que se vean otro tipo de sensores de posición. Las posibilidades que tenemos son, pues, bastante amplias:

a) **Potenciómetros:** esta ha sido una de las primeras posibilidades para la medida de posición angular ya que el dispositivo constituye —en sí mismo— un divisor resistivo que puede alimentarse en continua y cuya salida es proporcional al ángulo de giro. Su funcionamiento no puede ser más sen-

[3] Cualquier robot o brazo de robot puede llevar varios servomotores en función de sus posibilidades de movimiento (número de grados de libertad). En las cadenas de montaje —sobre todo en el campo de la automoción— los robots que realizan las tareas de montaje de vehículos, soldado de piezas, pintado, etc. giran en torno a varios ejes para ser capaces de alcanzar las posiciones adecuadas para llevar a cabo su tarea.

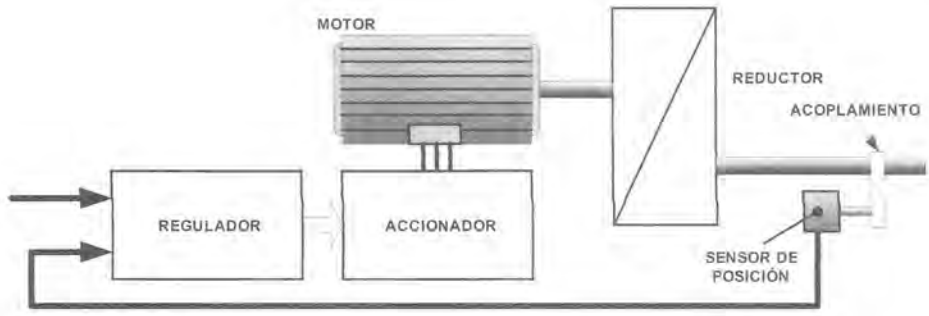


Figura 19.4. Diagrama de bloques de un servomotor de posición.

cillo ni tampoco los circuitos asociados. Sin embargo, los problemas que se generan por la necesidad de contacto físico entre el cursor y la superficie conductora del potenciómetro introducen factores de desgaste y par resistente que unidos al ángulo muerto hacen que el potenciómetro vaya siendo desplazado poco a poco del campo de medida de posición angular a favor de soluciones más *robustas*.

b) **Sensores electro-magnéticos:** son muchos los sensores que cabe introducir dentro de este apartado tales como el *syncro* (también conocidos como *seisin*) o el *resolver* [19.14]. Este tipo de sistemas presentan las ventajas propias de los sensores de tipo electromagnético, es decir, una gran robustez frente a sobrecargas y niveles de señal elevados con lo que tienen una buena capacidad de trabajo en ambientes con abundancia de interferencias. Sin embargo, presentan como inconvenientes su elevado peso, coste, su perturbación de la variable a medir y la presencia de contactos móviles lo que supone un desgaste y el consiguiente coste de mantenimiento.

c) **Sensores inductivos:** se pueden encontrar diversas versiones de este tipo de sensores con nombres también diversos e, incluso, es posible que al lector se le pueda ocurrir alguna combinación de bobinas, núcleos magnéticos y/o imanes que puedan servir para medir la posición angular empleando cualquiera de las configuraciones que hemos analizado en los capítulos anteriores o con topologías nuevas. Existen un par de sensores inductivos para la medida de posición angular bastante útiles ya que no tienen contactos móviles y de los que se pueden encontrar versiones comerciales; son el *magnesyn* y el *RVDT* [19.14].

El *magnesyn* consiste en un imán solidario con el eje cuya posición se quiere medir y un estator bobinado sobre un núcleo magnético toroidal de alta permeabilidad magnética con tres tomas situadas a 60, 180 y 300° (figura 19.5).

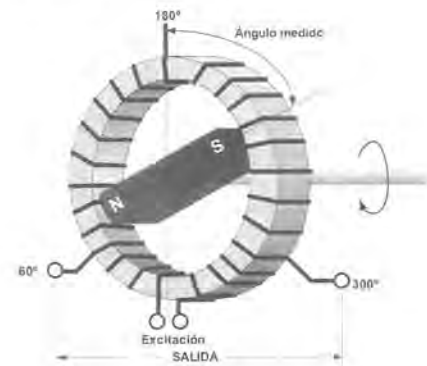


Figura 19.5. Magnesyn.

La tensión desarrollada entre las tomas de 60 y 300° es proporcional al coseno del ángulo del imán respecto del punto de 180°. Las ventajas e inconvenientes de este tipo de sensores son similares a los que presenta un sensor electromagnético por lo que sólo presenta frente a aquél la inexistencia de contactos móviles.

El RVDT (*Rotary Variable Differencial Transformer*) es similar al LVDT en funcionamiento aunque su forma geométrica es, lógicamente, diferente según puede observarse en la figura 19.6.

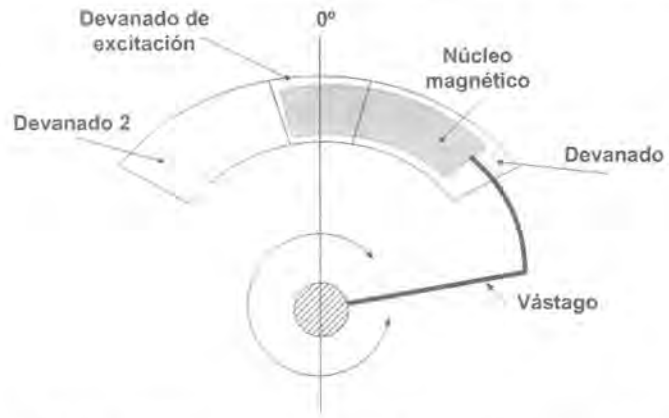


Figura 19.6. El RVDT es similar al LVDT excepto en su morfología.

La perturbación de la variable a medir es menor que en el caso anterior y su comportamiento tiene las ventajas propias de los LVDTs en cuanto a precisión y sensibilidad, existiendo circuitos ya preparados para gestionar su salida; su principal inconveniente hay que buscarlo en la limitación del ángulo de giro que no cubre más que un pequeño arco.

La lista de sistemas inductivos comerciales para medir posición es muy larga y son frecuentes nombres producidos por cada fabricante como el NACPS® (*Non-Contacting Angular Position Sensor*) [19.4] que basa su funcionamiento en la medida del desfase entre dos bobinas y que tampoco introduce demasiada perturbación sobre la medida.

d) **Codificadores de pulsos:** bajo este nombre o el anglosajón, *encoders*, se incluyen un buen número de sistemas que proporcionan una salida digital que informa sobre la posición de un determinado eje. A pesar de que su aspecto exterior e, incluso, su comportamiento pueda ser muy parecido, nos podemos encontrar con sistemas muy diversos y de tecnologías también diferentes.

Los *encoders* se pueden clasificar en función de múltiples criterios. Si el criterio es la tecnología, tenemos:

- Sistemas basados en efecto Hall que producen pulsos cada vez que un imán pasa por delante del sensor Hall.
- Sistemas inductivos que utilizan un detector de proximidad para contar los dientes metálicos de una rueda dentada solidaria al eje.
- Sistemas ópticos en los que un haz de luz (entre un LED y un fotodiodo) es cortado por una rueda dentada o con agujeros o en los que se produce o no reflexión sobre un disco sobre el que hay dibujadas líneas claras y oscuras.

Pero, si el criterio es el tipo de salida, se produce la clasificación más interesante desde el punto de vista de su aplicación:

- **Sensores absolutos:** son aquellos en los que la posición determina unívocamente la salida, es decir, en una misma posición siempre se tiene la misma salida. La inmensa mayoría de los sensores absolutos son ópticos por reflexión sobre un disco en el que se dibuja el código Gray y, aunque no tienen una resolución muy grande debido al espacio que ocupan emisores y receptores así como el propio código (salvo que el disco sea muy grande), pueden ser útiles en aplicaciones sencillas.
- **Sensores relativos** cuya salida informa del cambio de posición respecto de una posición de referencia. Estos pueden ser construidos en cualquiera de las tecnologías que antes se comentaron aunque la mayor parte de ellos son de tipo óptico. El funcionamiento básico sería muy sencillo ya que se cuentan pulsos a partir de una determinada posición y el total indica la posición. El sistema real debe ser algo más complejo ya que se debe poder contar y descontar según el sentido en que se desplace el eje; para ello se incluye una segunda hilera de detección desfasada con la primera para saber en qué sentido está girando[4].

A la hora de elegir un sensor de posición, la solución suele estar casi siempre en un dispositivo de tipo *encoder* y, en la mayoría de los casos, óptico ya que este tipo de sensores presenta costes bajos, pequeño volumen y peso (importante para no cargar innecesariamente el eje cuya velocidad se quiere medir) y un amplio abanico de oferta comercial.

Desde el punto de vista instrumental, la selección de un *encoder* para una aplicación concreta es muy sencilla ya que su salida es siempre de tipo digital (compatible TTL casi siempre) con lo que el árbol de decisión será como el que se muestra en la figura 19.7.

Figura 19.7. Criterios de decisión para la selección de sensores de posición angulares.

La capacidad de totalización de un *encoder* puede ser superior al número de pulsos totales que es capaz de producir en una vuelta; en este caso, el dispositivo será capaz de totalizar giros superiores a los 360°, es decir, será un dispositivo multivuelta. Otra opción disponible es el empleo de dispositivos de medida de

[4] En los dispositivos de este tipo se suelen proporcionar, además, las salidas que corresponden a los pulsos generados lo que se emplea para medir velocidad (usando su frecuencia). Esto es muy útil en los servomotores ya que, muchos de ellos, tienen una realimentación de velocidad de giro como en el caso de los motores de tipo asíncrono lo que les permite una sensible mejoría en el control.

velocidad —sólo producen pulsos— y realizar la totalización de la velocidad en circuitería externa, bien sea de lógica cableada o programada o bien basado en el empleo de algún tipo de microcontrolador.

Posición lineal

El problema de la medida de la posición lineal no trata de fijar las tres coordenadas de posición espacial en un volumen más o menos restringido, sino que se tratará de algo más simple: el cálculo de la distancia de un objeto determinado a un punto en el que situaremos un sensor. Lógicamente si el número de puntos desde los que se mide es de dos, se puede resolver la posición en un plano dado mientras que si es de tres, se puede obtener la posición espacial genérica según se muestra en la figura 19.8.

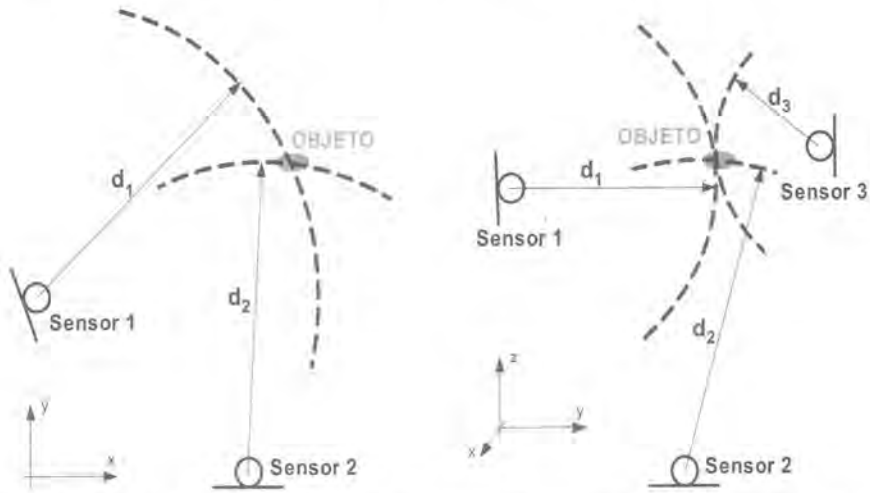


Figura 19.8. Obtención de la distancia a un objeto mediante dos sensores (en un plano) o tres sensores (en el espacio en general).

Lógicamente, la solución de posición deberá hacerse por algún método de triangulación y se podrá obtener más de una posición como solución por lo que será necesario incluir alguna restricción más al problema como la de considerar que el objeto sólo está en uno de los dos semiplanos o semiespacios. Por tanto, como antes dijimos, la resolución del problema de la medida de la posición se puede restringir a un proceso algo más simple: la medida de distancia a un objeto desde un punto[5].

La medida de la distancia a un punto se puede dividir en dos casos principales: medida sin contacto físico y medida con contacto físico; el primer caso, suele ser el menos habitual dentro del mundo de la instrumentación industrial mientras que el segundo es algo más frecuente.

La medida de la distancia a un objeto sin contacto físico se puede realizar de varias formas; para distancias medias (algunas decenas de metros) se pueden usar dispositivos que utilizan la técnica impulso-eco basados en ultrasonidos o por triangulación mediante luz [19.1], [19.13] generalmente láser aunque es posible emplear técnicas similares para distancias muy grandes en aplicaciones militares o de aviónica [19.29]. El funcionamiento de los telémetros láser consiste en la proyección de un punto sobre el objeto cuya distancia se quiere medir y su detección mediante un sensor de luz capaz de identificar la posición del punto como en el caso de un CCD lineal (figura 19.9).

[5] La posición en el espacio también se puede medir con métodos directos que funcionan como los radares o los sonares en los que no siempre se precisa la triangulación.

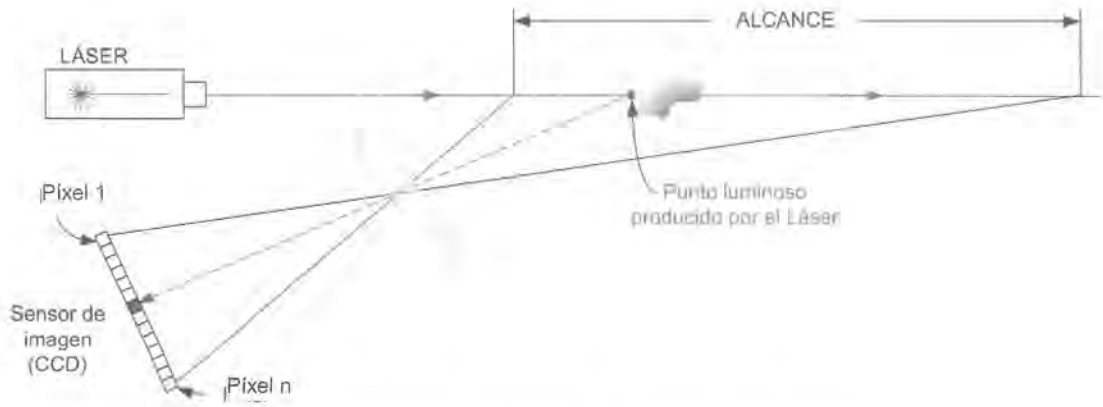


Figura 19.9. Medida de distancia sin contacto mediante láser.

Este tipo de sensores y otros similares se usan en múltiples campos que van desde complejas aplicaciones de navegación para barcos y aviones, identificación de objetivos en sistemas de bombas guiadas o sistemas de enfoque de cámaras de vídeo o fotográficas[6].

Hablando de distancias muy pequeñas (del orden de milímetros) los sistemas de medida pueden ser de carácter inductivo (alcances que oscilan entre 2 y 100 mm) que miden la influencia de un objeto ferromagnético en el campo generado por el dispositivo. Su linealidad no es muy buena aunque la necesidad de un material ferromagnético para detectar les hace muy selectivos. Menores alcances tienen los basados en corrientes de Eddy generadas sobre un material metálico en sus dominios magnéticos al igual que ocurre en los sensores de proximidad inductivos ECKO. La precisión de estos sistemas es baja y presentan algunos problemas tales como:

- Las propiedades del material influyen fuertemente en la calibración.
- La temperatura del material a detectar afecta a la medida.
- Es muy sensible a la presencia de campos magnéticos cercanos.

Alcances similares a los inductivos basados en las corrientes de Eddy tienen los capacitivos con márgenes que pueden ser mucho menores del milímetro. En estas cortas distancias su precisión es elevada para medir la distancia a objetos metálicos, siendo muy estables y no siendo afectada la medida por campos magnéticos ni por la temperatura.

Medida de distancia con contacto físico (medida de desplazamiento)

Este es el caso más frecuente de medida de distancia en el campo de la instrumentación industrial y se usa en diversos casos tales como:

a) Robótica.
b) Sistemas para realimentar la posición de un accionamiento.
c) Control de máquinas herramientas.
d) Medida del alargamiento de cables de sujeción.

A estos casos muy generales hay que añadir un extenso etcétera de casos más concretos. En general, en todos ellos, se dispone de un vástago cuyo desplazamiento (de ahí el nombre de la medida) actúa sobre el

[6] En algunas cámaras la detección de la distancia de enfoque se hace por otros métodos, tales como la maximización del contraste de la imagen empleando un CCD —en las cámaras digitales es el mismo que se emplea para tomar la imagen— con lo que se evita el problema derivado de la realización de fotografías a través de cristales.

sensor que proporcionará el valor de la medida final. Para efectuar esta medida se pueden usar diversas posibilidades:

- Potenciómetros lineales.
- LVDTs.
- *Encoders* lineales.
- Sensores de desplazamiento angular accionados por cable.

La mayoría de los sistemas de medida de desplazamiento por contacto tienen un alcance relativamente pequeño ya que este tipo de medida supone la presencia implícita de un vástago que debe tener una longitud igual al máximo valor de la medida. Por tanto, estos sensores rara vez sobrepasan el alcance de un metro. Sin embargo, en algunas ocasiones es preciso detectar el desplazamiento de un objeto mucho más allá de esos valores. En estos casos se puede emplear un *encoder* lineal de tipo relativo —muchas veces construido especialmente para la aplicación— cuyo alcance se puede considerar sólo limitado por la capacidad de los contadores que totalizan los pulsos (figura 19.10).

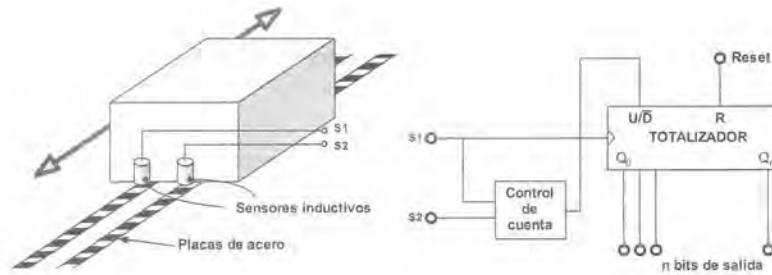


Figura 19.10. *Encoder* relativo construido a medida para medir la distancia recorrida por una máquina. El funcionamiento requiere dos sensores diferentes, inductivos, por ejemplo, para generar la orden de contar o descontar según el sentido del movimiento. Lógicamente, el totalizador será un contador que deberá permitir la cuenta ascendente y la descendente en función de la entrada U/D.

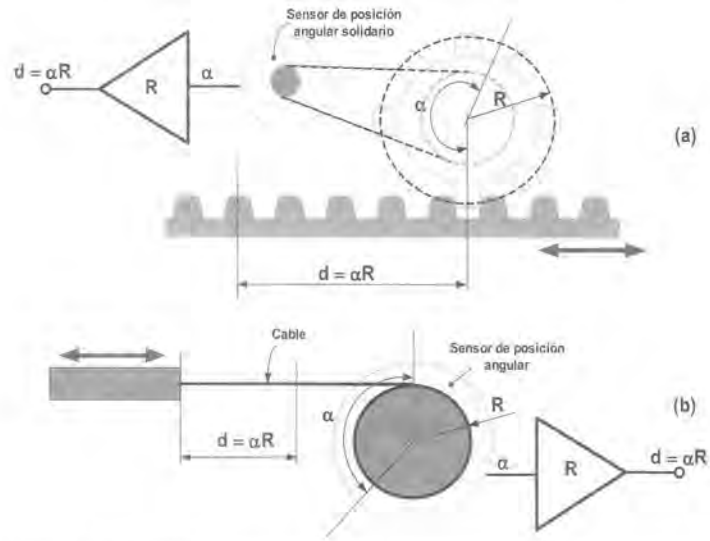


Figura 19.11. Medida de posición lineal mediante sensores de posición angulares: (a) mediante el aprovechamiento del propio sistema de movimiento; (b) mediante un accionamiento por cable.

También es posible emplear sensores de posición angular para la detección de la posición lineal, bien sea porque hay una relación clara entre ambas, como en el caso de un acoplamiento sin-fin corona (figura 19.11a) o mediante la inclusión de un vínculo mecánico como en el caso de los sensores angulares accionados por cable (figura 19.11b).

En la figura 19.12 se muestra un resumen de las posibilidades de medida de distancia lineal con diversos sensores.

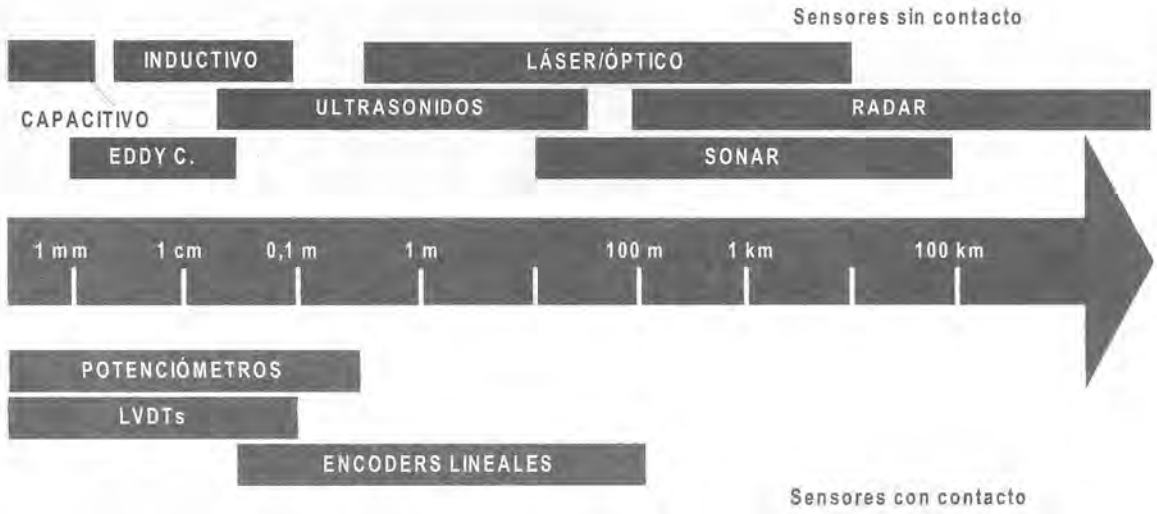


Figura 19.12. Posibilidades de medida de distancia con diversos sensores y su alcance.

19.2.3 MEDIDA DE LA VELOCIDAD

Si aseveramos que la velocidad de un objeto no es una medida frecuente, a casi todo el mundo le viene a la cabeza alguno de los controles de velocidad que las fuerzas de orden público realizan en nuestras carreteras y quizá alguno haya sufrido los efectos de las buenas características de los sistemas de medida empleados.

Sin embargo, este tipo de sensores no suele ser muy habitual excepto en vehículos y, aun en la mayoría de estos casos, la variable que se mide no es la velocidad del objeto sino algún parámetro que esté relacionado directamente con ella. Por contra, la velocidad angular sí que es una variable que se mide con gran frecuencia y, en algunos casos, a partir de ella se obtiene la velocidad lineal cuando el eje sobre el que se mide la velocidad angular mueve una rueda que impulsa a un vehículo. Además, en los controles de velocidad de motores de cualquier tipo, la velocidad angular es la salida del sistema por lo que deberá ser la variable realimentada.

A la hora de medir la velocidad angular existen, en principio, muchas opciones de sensores aunque la práctica diaria no aconseja el uso nada más que de unos pocos. Así, tenemos que los sensores de tipo electromecánico tales como tacogeneradores y dinamos tacométricas resultan bastante obsoletos y, aunque bien es cierto que se pueden ver algunos aún, su uso ha ido haciéndose menos frecuente hasta casi poder considerarse como nulo. Sus únicas características destacables son las de robustez y fiabilidad aunque su coste, peso y malas características de salida les han relegado casi al olvido. Tampoco son ya usuales artilugios como los

llamados "reguladores de bolas" que no son más que sistemas inerciales en los que la altura de las bolas depende de la velocidad de giro[7].

Los sistemas más empleados hoy en día son los generadores de pulsos con diversas tecnologías de construcción y que producen una cantidad de pulsos en cada unidad de tiempo proporcional a la velocidad del eje al que son solidarios (figura 19.13).

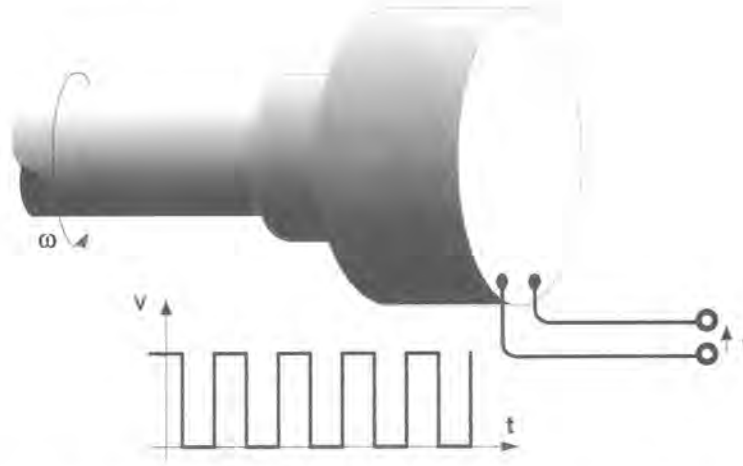
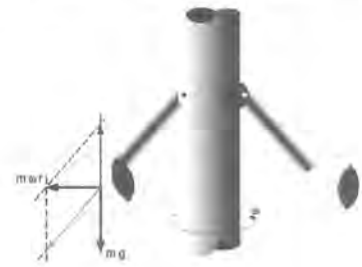


Figura 19.13. Generador de pulsos como sensor de velocidad y su salida.

Por lo tanto, los generadores de pulsos producen una señal cuya frecuencia es proporcional a la velocidad angular con lo que el tratamiento a realizar será digital, contando pulsos durante un cierto tiempo, al final del cual, tendremos en el contador un valor directamente proporcional a la velocidad angular que estamos midiendo. Así, resulta evidente que cuantos más pulsos se generen en cada unidad de tiempo, mayor va a ser el número que obtendremos en el contador y, consecuentemente, mayor va a ser la resolución de velocidad del sistema. Por otro lado, cuanto mayor sea el número de pulsos por unidad de tiempo que es capaz de proporcionar el sensor, mayor va tener que ser la capacidad de recuento (más bits) del contador.

En la mayoría de los sistemas, los ejes giran a velocidades relativamente lentas en relación con la velocidad de giro del motor que los mueve —igual que ocurre en un vehículo cuando va en las marchas bajas— ya que suelen llevar incorporado un reductor mecánico que incrementa el par a costa de reducir la velocidad de giro.

[7] Los reguladores de bolas que son el símbolo de la Ingeniería Industrial como titulación han sido de los primeros sistemas de regulación que la industria ha dispuesto y consistían en un par de bolas metálicas que giraban solidarias a un eje con capacidad de subir o de bajar según la velocidad de giro hasta que su peso era compensado por la fuerza centrífuga.



Si tratamos de medir la velocidad de giro del eje final (el de baja velocidad, que se suele denominar sólo el "eje de baja") tendremos valores menores de velocidad con lo que se producirán menos pulsos y la resolución del sistema será menor; por contra, si medimos el eje de alta velocidad (eje de alta), obtendremos para la misma velocidad de salida un tren de pulsos de mayor frecuencia con lo que conseguiremos una mayor resolución según se indica en la figura 19.14.

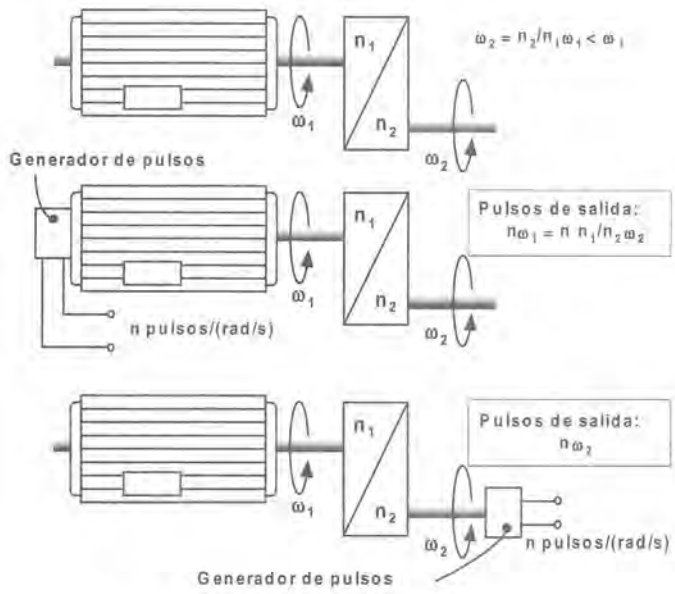


Figura 19.14. Posibles colocaciones de los sensores en los ejes de alta o baja (tras el reductor) de un motor para medir su velocidad de giro. La máxima resolución se obtiene cuando se sitúa en el eje de alta.

Lógicamente, la situación del generador de pulsos en el eje de alta no siempre es posible y dependerá de la disposición del sistema mecánico, de las características del motor, etc. con lo que, en algunos casos no queda más remedio que situarlo en el eje de baja. Aun en este caso, muchos de los generadores de pulsos de hoy en día poseen un número de pulsos de salida muy elevado incluso en condiciones de baja velocidad.

Los generadores de pulsos para la medida de la velocidad de ejes pueden emplear diversas tecnologías y, aunque los más populares son los ópticos, se pueden encontrar bastantes ejemplos de dispositivos de tipo inductivo o de efecto Hall.

Quizá el ejemplo más sencillo pueda ser alguno de los generadores de pulsos "caseros" empleados para medir la velocidad de una bicicleta que consisten en un sensor de efecto Hall de salida digital situado en la horquilla que sujeta la rueda y una serie de pequeños imanes dispuestos en los radios tal y como se indica en la figura 19.15.

Usando el mismo principio de funcionamiento, algunos fabricantes incorporan una rueda con polos magnéticos que producen pulsos al pasar al lado del sensor de efecto Hall, de tal manera que se consigue un efecto similar.

Otras soluciones tecnológicas de bajo coste y resolución limitada emplean una rueda dentada que intercepta un haz de luz entre un par LED-fotodiodo como ocurre en los ratones de PC, aunque los generadores

ópticos habituales son algo más complejos, funcionan por reflexión en lugar de por interceptación de haz y suelen incorporar dos salidas para proporcionar el sentido de giro.

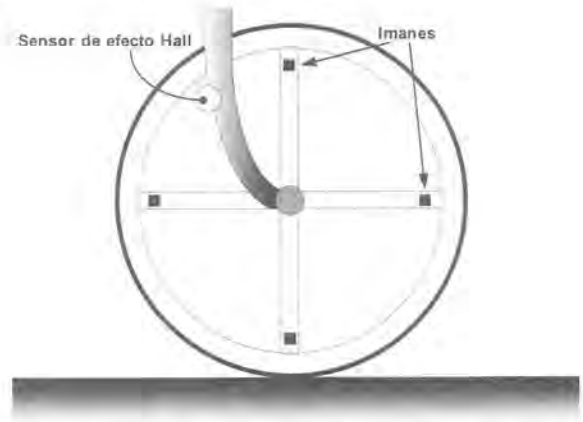


Figura 19.15. Medida de la velocidad de la rueda de una bicicleta empleando un dispositivo de efecto Hall y varios imanes situados sobre los radios.

19.2.4 MEDIDA DE LA ACELERACIÓN Y DE LA VIBRACIÓN

La medida de la aceleración ha alcanzado recientemente cotas de utilización muy altas, en parte gracias a las excelentes prestaciones de los sensores desarrollados para ser aplicados en sistemas de seguridad en automoción como en el caso del *airbag*. Los primeros sensores de aceleración y vibración eran auténticos sistemas de alta complejidad y reducida fiabilidad que se basaban en la medida de los desplazamientos de una masa inercial sujeta a la aceleración con resortes que contrarrestaban el efecto de la fuerza generada por la masa.

De la misma forma, la medida de la vibración resulta muy parecida a la de la aceleración ya que la presencia de movimientos vibratorios supone aceleraciones cambiantes cuyo aspecto es similar al del movimiento inicial.

Otras variables habituales que llevan implícita la medida de la aceleración que da lugar a los **sensores de impacto** que se caracterizan por la detección de fuertes aceleraciones en cortos periodos de tiempo como en el caso de los sensores de choque que disparan los *airbag*.

En todos estos casos cabe hablar de medidas de aceleración y los sensores para cada aplicación serán siempre acelerómetros aunque, en función de lo que se vaya a medir, el dispositivo estará diseñado para ofrecer las mejores características en ese sentido: por ejemplo, cuando se van a medir aceleraciones constantes no tiene sentido el empleo de dispositivos piezoeléctricos puesto que no responden bien ante solicitaciones constantes. Cuando se pretende medir una vibración, muchas veces sólo interesa su amplitud y/o su frecuencia con lo que la linealidad del dispositivo que la mida es un factor secundario. Así pues, aunque la medida de aceleración genérica tenga varias vertientes, en muchas ocasiones los sensores se denominan como sensores de vibración, de choque o impacto a pesar de que todos ellos puedan ser agrupados dentro del campo de los acelerómetros porque, de hecho, lo son. Finalmente, cabe considerar los sensores de golpe[8] (*tilt sensors*) que son un caso de sensor de impacto para detectar pequeños golpes.

[8] Seguramente, el lector conocerá el término *tilt* (falta) de las máquinas de juego tipo *pin-ball* que detectaban los movimientos excesivos mientras intentábamos evitar que la bola "se nos colase". En esos casos, se empleaba también un acelerómetro muy básico y con salida digital formado por un péndulo y un aro, ambos metálicos: mientras el péndulo no tocase el aro, todo iba bien pero, si, por un golpe, el péndulo tocaba el aro los mandos se bloqueaban y veíamos cómo la bola se colaba irremisiblemente.

Los métodos para medir aceleración son muy variados y aquí sólo vamos a mencionar algunos de ellos:

a) **Acelerómetros mecánicos** que emplean una masa inerte y resortes elásticos y en los que los cambios se miden con galgas extensométricas, incluyendo sistemas de amortiguación que evitan la propia oscilación. Otros sistemas emplean sistemas rotativos desequilibrados que originan movimientos oscilatorios cuando están sometidos a aceleración (*servoacelerómetros*) o detectan el desplazamiento de una masa inerte mediante cambios en la transferencia de calor (*acelerómetros térmicos*).

b) **Acelerómetros capacitivos**; modifican la posición relativa de las placas de un microcondensador cuando está sometido a aceleración.

c) **Acelerómetros piezoeléctricos**; su funcionamiento se basa en el efecto piezoeléctrico y son, probablemente, de los más usados en la medida de vibraciones. Su principal inconveniente radica en su frecuencia máxima de trabajo y en la incapacidad de mantener un nivel permanente de salida ante una entrada continua.

d) **Acelerómetros micromecánicos** (MEMS según las siglas anglosajonas *Micro-Electro-Mechanical System)*; este tipo de dispositivos ha sido desarrollado para su empleo como sensor de impacto en los sistemas de *airbag*, en sistemas antibloqueo de frenos o en cualquier otro proceso en que se pretenda medir impacto.

Para elegir un sensor para una medida de aceleración de entre las posibilidades del mercado, además de los márgenes de valores de la aceleración que admite habrá que tener en cuenta si es capaz de medir en continua o sólo en alterna, la máxima frecuencia a la que puede trabajar, así como los correspondientes parámetros instrumentales típicos de todo sensor. En la tabla 19.7 se resumen algunas de las principales características de los acelerómetros y sus aplicaciones más típicas teniendo en cuenta que el margen de medida se expresa en unidades g (aceleración de la gravedad terrestre cuyo valor es aproximadamente[9] de 9,81 m/s^2) [19.10].

Tipo	Margen de medida	Ancho de banda (Hz)	Ventajas e inconvenientes	Aplicaciones
MEMS	1,5 a 250 g	0,1 a 1500	– Alta sensibilidad (hasta 1,3V/g) – Coste medio – Uso sencillo – Bajas temperaturas	– Impacto – ABS – Airbag – Uso en automoción
Piezo-eléctricos	0 a 2000 g	10 a 20000	– Sensibilidad media – Uso complejo – Bajas temperaturas – No funcionan en continua	– Vibración – Impacto – Uso industrial
Piezo-resistivos	0 a 2000 g	0 a 10000	– Respuesta en continua y alterna – Prestaciones medias – Bajo coste	– Vibración – Impacto – Automoción
Capacitivos	0 a 1000 g	0 a 2000	– Funciona en continua – Bajo ruido – Baja potencia – Excelentes características	– Uso general – Uso industrial
Mecánicos	0 a 200 g	0 a 1000	– Alta precisión en continua – Lentos – Alto coste	– Navegación inercial – Guía de misiles – Herramientas – Nivelación

Tabla 19.7. Algunos tipos de acelerómetros.

[9] El valor de la aceleración de la gravedad es variable debido a múltiples circunstancias como la no esfericidad de la tierra y la distinta velocidad lineal de los puntos situados en diversas latitudes. Los valores exactos deberán buscarse para cada lugar.

En muchas ocasiones, no sólo es necesario medir la aceleración en un eje, sino que se precisa efectuar la medida en dos o tres; pues bien, los fabricantes de acelerómetros proporcionan soluciones biaxiales y triaxiales capaces de proporcionar el valor de aceleración en varios ejes ortogonales.

Una de las aplicaciones de los acelerómetros donde se precisa la medida en dos ejes es la medida de la inclinación, variable que resulta muy útil en algunos casos tales como la monitorización del comportamiento de vehículos o los sistemas de nivelación de suelo empleados en algunas máquinas. En estos casos, se están midiendo las dos componentes en que se descompone la aceleración de la gravedad según dos ejes con lo que es factible conocer la inclinación según se indica en la figura 19.16.

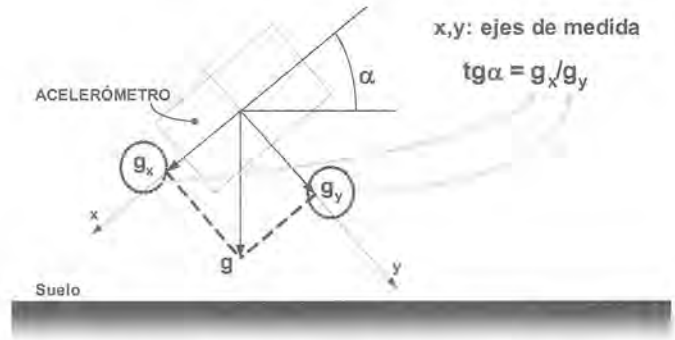


Figura 19.16. Utilización de un acelerómetro de dos ejes para la medida de la inclinación.

En cualquier caso, la selección del acelerómetro para una aplicación concreta se hará en función de una serie de criterios:

a) Frecuencia de trabajo o margen de frecuencias de uso, tanto los valores mínimos —¿podrían trabajar en continua? — como los máximos que determinan la velocidad de respuesta que precisamos.

b) Los valores máximos y mínimos del nivel de la señal que esperamos. El valor mínimo de señal no suele ser muy importante excepto para algunas aplicaciones en concreto en que se precisan medidas de vibraciones muy débiles. En estos casos, el valor de la señal debería ser de, al menos, cinco veces el valor del ruido generado [19.31]. Por el contrario, el valor máximo no debería sobrecargar el sensor ni introducir distorsiones en la señal (tanto el sensor como el amplificador podrían salirse de la zona lineal).

c) Consideraciones acerca de la forma de montaje, el espacio disponible, la forma de salida de los cables, etc. pueden parecer triviales al usuario, pero pueden hacer que una selección quede invalidada.

d) Otras consideraciones tales como la temperatura de trabajo, aspectos ambientales y de compatibilidad química o la necesidad de seguridad intrínseca deberán tenerse en cuenta a la hora de efectuar la elección final del sensor.

19.2.5 MEDIDA DE FUERZA Y DE PESO

Las medidas de fuerza y de peso son bastante habituales desde hace mucho tiempo ya que se han venido usando para cuantificar la masa de los productos en las transacciones comerciales; aún hoy en día se pueden ver diversos tipos de "romanas" empleadas en algunos mercadillos o balanzas de dos brazos con complejos sistemas de amortiguación y elevadísima precisión guardadas en urnas de cristal para protegerlas o para evitar que nada interfiera en la medida. La evolución de la tecnología hizo que los tiempos de las balanzas

que empleaban la compensación de fuerzas dieran paso a las dinamométricas que emplean un resorte como fuerza que contrarresta el peso y, también estas han llegado hasta nuestros días en diversas formas, algunas de las cuales se han convertido en tan impopulares como la báscula de baño que siempre insiste en recordarnos dónde acaban los excesos alimentarios.

La diferencia entre sensores de fuerza y sensores de peso es sólo debida a las "especiales" condiciones que impone el peso, fuerza asociada a una masa y que sólo se puede aplicar en una dirección y sentido: vertical hacia abajo. Los sensores de peso tendrán, en general una base sólida sobre la que apoyarse y aplicarán la masa que se desea pesar justo encima. Por el contrario, los sensores de fuerza genéricos, deberán ser capaces de medir la fuerza en cualquier dirección y sentido por lo que las configuraciones mecánicas estarán más abiertas y serán bastante diferentes en función del caso.

Pero, en cualquiera de las dos situaciones, el principio de medida será muy similar excepto por las consideraciones dinámicas que se impongan: el peso siempre es una fuerza estática (no cambia con el tiempo) por lo que los sensores no tendrán restricciones frecuenciales; por el contrario, deberán proporcionar una correcta medida en continua.

El peso es una medida de fuerza de tipo estático, es decir, una medida en la que la fuerza ocasionada por una determinada masa al ser atraída por la fuerza de la gravedad permanece constante o cuasi-constante en el tiempo. Por tanto, la medida de peso no deja de ser más que una medida más de fuerza y, como tal, todos los sensores destinados a medir fuerzas se pueden destinar a medir pesos y, básicamente las precauciones y criterios a tener en cuenta son los mismos.

Sin embargo, dado lo frecuente de este tipo de medida en todos los niveles, desde una modesta (y molesta a veces) báscula de baño hasta los sistemas de pesaje de gran tonelaje para camiones o los de gran precisión para laboratorios químicos, se han desarrollado sensores especiales preparados para la medida de peso y que se denominan células o **celdas de carga** (*load cells*). Estos dispositivos son piezas mecánicas que se deforman de forma controlada con el peso aplicado sobre ellas y cuya deformación se mide mediante algún tipo de sensor electrónico según se muestra en la figura 19.17.

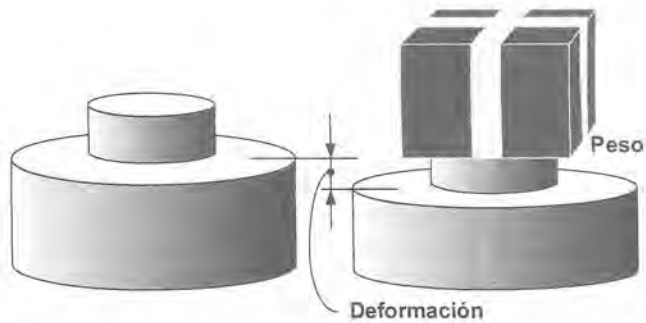


Figura 19.17. Uso de celdas de carga como sensores de peso.

El dispositivo mecánico se diseña para conseguir que tenga las mejores prestaciones posibles y que se adapte al lugar de la estructura mecánica donde va a ser colocado. En la figura 19.18 se muestran algunas realizaciones posibles de celdas de carga que son habituales en el mercado.

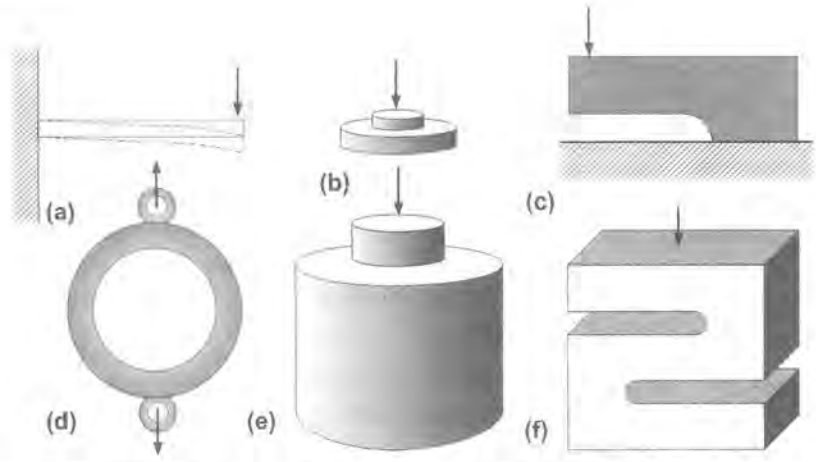


Figura 19.18. Diversos sensores de peso y fuerza.

Sobre cualquiera de estos dispositivos se sitúan los sensores electrónicos que producirán la señal. Estos sensores son de los siguientes tipos:

a) **Galgas metálicas:** que producen bajos niveles de señal pero poseen una linealidad bastante alta y una buena precisión. La necesidad de amplificación para producir niveles razonables de señal fue un inconveniente grave que dificultó su uso aunque, hoy en día, la bajada del coste de los sistemas de amplificación y la aparición de excelentes dispositivos monolíticos para amplificar las señales procedentes de puentes de medida ha vuelto a relanzar su uso para aprovechar sus buenas prestaciones.

b) **Galgas piezorresistivas:** similares a las anteriores pero construidas con material semiconductor, producen niveles mucho más elevados de señal no precisando, en la mayoría de los casos, ni siquiera un amplificador ya que los niveles que proporcionan son suficientes para muchas aplicaciones. En su contra tienen una baja linealidad y una cierta dependencia de la temperatura.

c) **De magnetostricción:** cambian su permeabilidad con el esfuerzo aplicado; no es un sistema muy habitual pero sigue usándose en rodillos de laminación en frío o en caliente.

Los dispositivos anteriores se sitúan en la estructura del sensor y traducen el esfuerzo de deformación en un cambio de valor de su resistencia; el montaje suele ser en puente completo y con dos galgas trabajando a tracción y dos a compresión con lo que la sensibilidad se multiplica por cuatro respecto a la que posee una sola galga. De los dispositivos de la figura 19.18, los que trabajan a flexión (figura 19.18a) son los más simples —muchas veces la construcción es "casera"— presentan un coste muy bajo y las galgas se sitúan sobre la superficie implicando la necesidad de protección frente al ambiente. Los que son de tipo ménsula (figura 19.18c) son similares a los anteriores pero las galgas se sitúan en el interior y dada su mayor robustez son capaces de soportar esfuerzos laterales sin afectar a la medida. Los de tipo botón (figura 19.18b) son sencillos, baratos y tienen un margen de medida bajo para aplicaciones sencillas como las de una báscula de baño. Sus hermanas mayores (figura 19.18e) constituyen la más habitual de las realizaciones de células de carga y suelen estar previstas para trabajar con valores de carga muy grandes (varias toneladas). Las realizaciones de las figuras 19.18d y 19.18f permiten el trabajo a tracción y, en el último caso, también a compresión.

A la hora de seleccionar un sensor para la medida de peso habrá que tener en cuenta el margen de medida que determinará, en una buena parte, el dispositivo final elegido. Sin embargo, aspectos tales como el margen de temperaturas de uso son muy importantes —sobre todo con sistemas que incorporen galgas piezorresistivas— ya que pueden afectar seriamente a la precisión del conjunto.

Pero hay un aspecto que debe tenerse en cuenta en el caso de las celdas de carga y es la capacidad de sobrecarga. Esto no significa que la celda sea capaz de proporcionar una medida válida fuera de su margen de uso, sino que es capaz de soportar mecánicamente una sobrecarga sin que quede inutilizada. En este punto hay que recordar que la inutilización o daño permanente puede ocurrir si se supera el límite elástico del material de la celda y se entra en la zona de deformación plástica. En la figura 19.19 se muestra una gráfica que indica el significado de la sobrecarga en relación con los márgenes de medida y con los límites elásticos de los materiales.

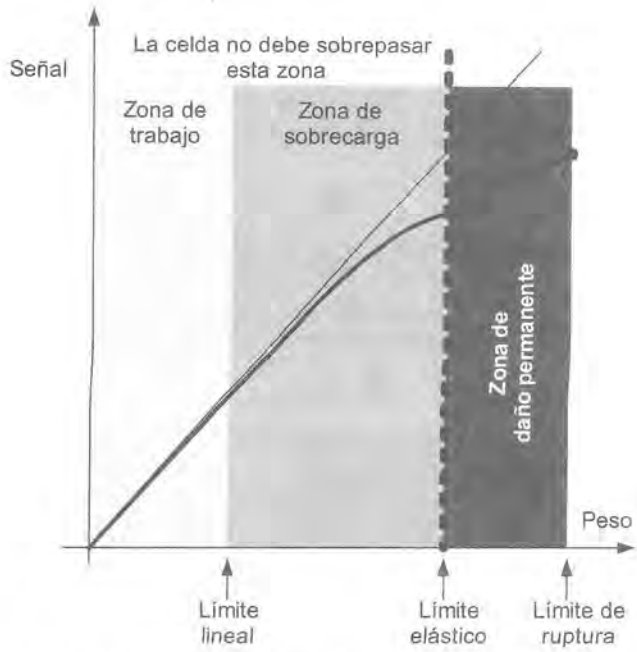


Figura 19.19. Límite de uso de las celdas de carga. No debe trabajarse en la zona más a la izquierda ya que el daño es irreparable.

Normalmente las celdas de carga suelen permitir sobrecargas elevadas, del 50 o del 100%, incluso bastante más.

Otro factor importante es el de los tipos de esfuerzos que soportan: una celda de carga debe proporcionar una señal correspondiente al peso que se coloca sobre ella; sin embargo, la presencia de esfuerzos laterales puede falsear esa medida si la celda no es capaz de absorberlos sin modificar la lectura. Cuando se elige una celda de carga habrá que tener estos factores en cuenta si van a presentarse estas situaciones en su funcionamiento.

19.2.6 MEDIDA DE LA PRESIÓN

Tipos de medidas de presión

La medida de presión puede realizarse de diversas formas en función de la variable que realmente se desee cuantificar:

La primera idea es la medida de la **presión absoluta**, es decir, el valor de presión en un determinado punto de un medio respecto del cero absoluto de presión que supone el vacío perfecto. Una medida de este tipo es la de la presión atmosférica que determina, para un determinado lugar, el valor de esta magnitud respecto del vacío perfecto.

Sin embargo, en otras ocasiones, el valor absoluto de la presión no presenta demasiado interés, sino que lo que importa es la diferencia de presión entre dos puntos determinados; en este caso, la variable que se mide es la **presión diferencial**. Un caso muy típico es el de la pérdida de presión en una línea de fluido que circule por una tubería.

La tercera forma de medir presión es muy similar a la anterior, pero cuando uno de los puntos tiene la presión atmosférica: se trata de la **presión relativa** (*gauge pressure* o *gage pressure*) que es la diferencia de presión que existe entre un determinado punto y la atmosférica. Quizá la medida más conocida de presión relativa es la medida de la presión sanguínea (la popular tensión).

En la figura 19.20 se muestra un gráfico que resume el sentido de estas diversas medidas de presión. Lógicamente, cuando se va a elegir un sensor, lo primero que debemos hacer es decidir cuál de los tres tipos de medida queremos hacer y, en función de ello, decidir el tipo de sensor que precisamos.

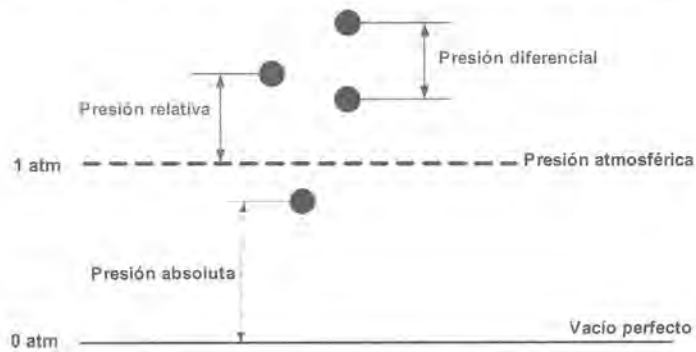


Figura 19.20. Sentido de las diversas medidas de presión en relación con la presión atmosférica o con el vacío perfecto.

Otra de las formas de medir presión es la medida de **vacío**; el vacío es una presión diferencial negativa respecto a la atmosférica y aunque los sensores que lo miden suelen dar la indicación como positiva. En estos casos hay que añadir la indicación "de vacío" a la medida para saber de qué estamos hablando y no dar lugar a equívocos. Por ejemplo, si un sensor de vacío conectado a una línea proporciona una medida de 50 kPa significa que la presión de la línea es –50 kPa respecto de la atmosférica, –50 kPa de presión relativa o, lo que es lo mismo, 50 kPa de vacío.

Aunque el lector seguramente lo conocerá, vamos a recordar cuáles son las unidades más típicas de presión que se usan y sus equivalencias y que no se restringen a las del Sistema Internacional de unidades. Es frecuente encontrarnos con equipos y sensores que proporcionan los datos como **psi** (*Pounds per Square Inch*, es decir, libras por pulgada cuadrada) típica en los países anglosajones, pero cada vez menos habitual, **atm** (atmósferas) que establece como valor unitario el que corresponde a la presión atmosférica a 0 m de altura, **mm de Hg** o **torr** (muy tradicional y rememorando el experimento de Torricelli) o, incluso, **cm de H_2O**, o el **bar** (más bien el **mbar** que es muy usado por los meteorólogos para hablar del anticiclón de las Azores o de la borrasca de las islas británicas). En la tabla 19.8 se proporcionan estas equivalencias aunque, en lo que sigue, usaremos la unidad del S.I. —el **Pascal (Pa)** equivalente a 1 N/m^2— como estándar aunque, debido a su bajo valor, en uno de sus múltiplos, el **kPa** que equivale a 1 mbar.

1 kPa	1	0,14504	0,0098694	689,46	10,198	10^{-3}
1 psi	6,8946	1	0,068046	51,714	70,310	0,068496
1 atm	101,32	14,696	1	759,83	1033,3	1,0066
1 mm Hg	0,13332	0,019337	0,0013161	1	1,3596	0,0013248
1 cm H_2O	0,098062	0,014223	0,00096780	0,73551	1	0.00097416
1 bar	10^3	14,504	0,99343	754,85	1026,5	1
	kPa	psi	Atm	mm Hg	cm H_2O	bar

Tabla 19.8. Equivalencias entre unidades de presión.

Sensores para la medida de presión

Casi la totalidad de los sensores para la medida de presión emplean algún tipo de dispositivo mecánico que se deforma y cuya deformación se traslada al sensor propiamente dicho que será de alguno de los tipos estudiados anteriormente. Esto significa que se traduce la presión a una variable intermedia más sencilla de medir mediante algún tipo de sensor primario como los que se muestran en la figura 19.21.

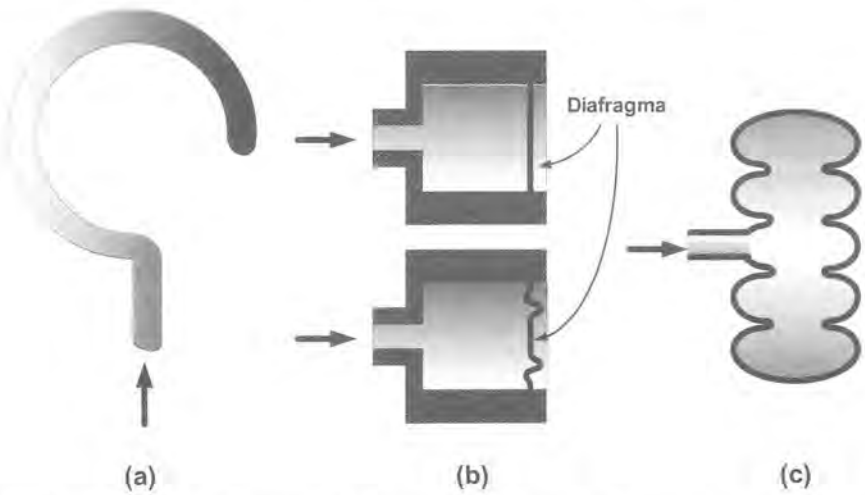


Figura 19.21. Sensores de presión que transforman el valor de presión en una deformación: *(a) tubo de Bourdon; (b) sensores de diafragma; (c) sensor de cápsula cerrada.*

El elemento sensible puede ser un tubo Bourdon (figura 19.21a), una membrana (figura 19.21b) o una cápsula (figura 19.21c) y actuará sobre el sensor eléctrico que será solidario a él según se indica en la figura 19.22.

Figura 19.22. Diagrama de bloques de un sensor de presión con paso por una variable intermedia.

El tipo de sensor eléctrico final determina la denominación del sensor de presión con lo que tendremos sensores potenciométricos, inductivos, capacitivos, piezoeléctricos, de galgas u ópticos con las ventajas e *inconvenientes inherentes a cada uno de estos tipos.* A pesar de que, en principio, podría utilizarse cualquier sensor primario en combinación con cualquier sensor eléctrico, lo cierto es que se tienen determinadas configuraciones habituales favorecidas por las facilidades constructivas. Seguidamente comentaremos brevemente cuáles son las ventajas e inconvenientes de cada tipo de sensor de presión.

- **Sensores potenciométricos:** la mayoría de ellos utilizan el tubo de Bourdon (figura 19.21a) cuyo extremo mueve el cursor de un potenciómetro circular. La presencia de la histéresis mecánica propia de los potenciómetros junto con sus limitaciones hacen de este sensor un dispositivo de bajas prestaciones pero —eso sí— de muy bajo coste lo que lo faculta para aplicaciones sencillas. Los márgenes de medida con este tipo de dispositivos no son muy amplios y pueden andar entre 35 kPa y 70 MPa [19.22].

- **Sensores inductivos:** el conjunto está formado en la mayoría de los casos por un LVDT y un dispositivo de tipo diafragma o cápsula (figuras 19.21b y 19.21c). La principal ventaja hay que buscarla en la presencia del LVDT cuyas prestaciones de precisión son muy altas aunque presenta serias limitaciones frecuenciales y —debido a que se utiliza un núcleo magnético abierto— presenta también problemas de interferencias magnéticas. Los márgenes de uso son similares a los anteriores aunque la precisión que se puede alcanzar es mejor y ronda el 0,5% de error máximo.

- **Sensores capacitivos:** utilizando un diafragma metálico como placa-electrodo de un condensador y otra placa fija se puede construir un sensor de presión capacitivo de buenas prestaciones aunque con la limitación de que se ve afectado por las vibraciones. Esta limitación es muy seria cuando el sensor está previsto para trabajar con bajas presiones ya que la construcción del diafragma debe hacerse con material muy fino cuya sensibilidad es muy elevada. Este tipo de sensores constituye uno de los más habituales sensores diferenciales usando el diafragma para separar las dos zonas de presión y es capaz de medir desde presiones diferenciales nulas hasta los 70 MPa.

- **Sensores piezoeléctricos:** si el sensor eléctrico es un dispositivo piezoeléctrico, se tiene un sensor de presión piezoeléctrico capaz de funcionar de forma bidireccional, es decir, es capaz de producir una señal eléctrica cuando se le aplica una presión (un movimiento) o de ocasionar el movimiento (y generar presión) cuando se le excita con una señal eléctrica[10]. El principal problema de este tipo de sensores lo constituye la dificultad para trabajar con excitaciones constantes según se pudo comprobar en el capítulo dedicado a ellos.

- **Sensores de galgas metálicas:** situando adecuadamente las galgas sobre un diafragma se puede medir su deformación y, así, obtener la presión que ha causado esa deformación. Normalmente se emplean dos o cuatro galgas (en este último caso, dos trabajan a tracción y dos a compresión). Los sistemas así obtenidos poseen un bajo nivel de salida lo que implica la necesidad de amplificar la señal mediante un amplificador diferencial (normalmente una configuración de instrumentación) aunque tienen la ventaja de excelentes características de repetibilidad y de histéresis. Una gran ventaja de estos tipos de sensores está en el gran margen de temperaturas de trabajo que los hace especialmente útiles en muchas aplicaciones industriales en las que los gases cuya presión se quiere medir están a temperaturas de algunas centenas de grados. Los valores de presión que se pueden medir son bastante altos y se pueden llegar desde 0,7 kPa hasta 1400 Mpa [19.22].

- **Sensores de galgas de semiconductor:** están basados en el efecto piezorresistivo y aunque podrían usarse de forma similar al caso anterior, en este tipo de sensores, todo el diafragma suele estar cons-

[10] Los ejemplos más típicos de este tipo de dispositivo funcionando como bidireccional lo constituye un altavoz/micrófono piezoeléctrico o un emisor/receptor de ultrasonidos. En ambos casos se produce una onda de presión (sonido) cuando se excita con señales eléctricas o se generan señales eléctricas cuando recibe una onda de presión (sonido).

truido en silicio y las galgas están depositadas sobre él por técnicas microeléctrónicas. El resultado es un diafragma con sensores integrados que posee una excelente sensibilidad y es capaz de producir valores elevados de salida lo que constituye una ventaja sobre los anteriores; otra importante ventaja radica en el propio material usado para hacer el diafragma que puede ser bastante inerte y que —por ello— tiene una excelente compatibilidad química con muchas sustancias. Los principales inconvenientes vienen de su elevado valor de histéresis, de la menor repetibilidad y del menor margen de temperatura de trabajo comparada con los sensores anteriores.

- **Sensores ópticos:** se basan en el cambio de los niveles de luz entre un emisor (normalmente un LED) y un receptor (normalmente un fotodiodo) cuando la presión modifica un elemento mecánico como los de la figura 19.21. El modo de funcionamiento puede ser por cambios en la reflexión o interrupción de haz. No suelen ser muy habituales todavía pero pueden alcanzar bastante precisión (errores menores del 0,1%) con grandes márgenes de trabajo (desde unos 35 kPa hasta más de 400 MPa).

Todos los anteriores sensores proporcionan niveles de señal cuando se someten a unas condiciones de presión. Sin embargo, también podemos encontrarnos con sensores cuyo objetivo es producir una señal todo-nada y que conmuta de una a otra situación en un determinado nivel de tensión: son los **interruptores de presión** que se pueden hacer como tales directamente añadiendo un contacto electromecánico a cualquiera de los sensores básicos de la figura 19.21 o bien se pueden desarrollar con un comparador a la salida de algún sensor de los antes comentados. Los interruptores de presión suelen emplearse en sistemas de seguridad de instalaciones para cortar el funcionamiento de una parte o todo un sistema si se superan determinados niveles de presión en algún punto que puedan poner en peligro el proceso o, incluso, la instalación o los usuarios.

Hay algunos procesos industriales que suponen el empleo de presiones muy elevadas y que requieren el empleo de equipos y —por supuesto— sensores especiales preparados para trabajar con los previsibles elevados niveles de presión. Los sistemas de alta presión y alto vacío son esencialmente iguales a los utilizados para menores presiones aunque los sensores básicos primarios están diseñados con una gran robustez mecánica para soportar los esfuerzos a que estarían sometidos. No hay una frontera clara entre lo que se considera una presión normal y lo que se denomina **alta presión**[11] aunque se suele considerar el valor de 69 bar [19.22] como el límite entre ambos casos.

Consideraciones en la selección y uso de sensores de presión

La selección de sensores de presión no parece compleja y basta con echar un vistazo a las características antes indicadas para hacerse una idea de que la mayoría de los sistemas de medida tienen un carácter más o menos universal por lo que no serían imprescindibles más consideraciones generales.

Por otro lado, los sensores de presión se comercializan como tales o incluyendo los correspondientes circuitos de adaptación y proporcionando una salida de alto nivel. En el primer caso, se tienen valores bajos de salida (algunas centenas de milivoltios), mientras que, en el segundo, producen valores estándar de salida, bien sea de corriente o de tensión; en este caso, estamos en presencia de un **transmisor de presión**.

A pesar de esta aparente sencillez de la elección, habrá que tener precauciones para no cometer errores absurdos a la hora de seleccionar; así, tendremos que tener en cuenta:

a) La **compatibilidad química** del medio: en muchos casos, las medidas de presión se hacen en medios que pueden ser muy agresivos para el sensor y que podrían suponer su degradación progresiva o, incluso, su destrucción casi inmediata. El sensor debe ser elegido acorde con el medio sobre el

[11] En algunos procesos como el de la producción de diamantes industriales se emplean presiones muy elevadas que llegan a rozar los 7000 bar. También en aplicaciones marinas para altas profundidades los equipos de medida de presión deben soportar valores muy elevados (se llegan a alcanzar muchos miles de bar de presión).

que se va a medir para evitar su mal funcionamiento o, en el caso más extremo, poner en peligro la integridad de la instalación o de las personas[12].

b) La capacidad de **soportar sobrepresiones**: aunque el margen de temperatura sea uno de los más importantes criterios a la hora de seleccionar un sensor, la capacidad de soportar sobrepresiones es un parámetro de seguridad igual de importante. Establece la capacidad del sensor de soportar una presión superior al margen de medida sin que sea destruido o dañado y, en la mayoría de los casos debe tenerse en cuenta ya que se dan muchas circunstancias en las que las líneas tienen presiones muy superiores a las previstas.

   Podría, claro está, elegir para las peores condiciones pero eso no resulta conveniente: si queremos medir entre 0 y 100 bar, ocurre una sobrepresión del 100% una vez cada mes —por ejemplo— y se elige un sensor con margen de medida de 200 bar, casi todo el tiempo estaría desaprovechado y sólo una vez al mes se ocuparía el extremo de medida (eso sí, mediríamos con exactitud la sobrepresión). Sería mejor solución emplear un sensor con el margen adecuado (100 bar) pero que sea capaz de soportar (no de medir correctamente) la sobrepresión.

c) El **margen de temperatura** de trabajo: la temperatura de trabajo del sensor afecta a la medida por lo que deberá ser tenida en cuenta bien sea en el propio sensor (hay sensores compensados térmicamente) o bien en el circuito que recibe la señal del sensor. En cualquier caso, fuera de los márgenes de temperatura o la medida es incorrecta o, incluso, el sensor puede ser dañado. Para trabajar en condiciones de alta temperatura hay sensores especiales (los más adecuados son los basados en galgas metálicas) pero en algunos casos se recurre al enfriamiento del sensor mediante células Peltier.

El uso de los sensores de presión es tan sencillo como cualquier otro aunque las particularidades de la variable a medir hacen que no se puedan colocar o retirar de una instalación de cualquier forma: debido a que el sensor tiene que tener acceso a la línea en la que se está midiendo, a la hora de efectuar la conexión del sensor la instalación o su retirada de la misma, esta debe estar sin presión —o vacío— para evitar fugas del medio generando un peligro de accidente potencial. Esto significaría que, en la mayoría de los casos, la instalación debería estar parada. Otra opción sería la de utilizar una derivación de la línea principal de medida con dos válvulas que permitan "aislar" la zona en la que se situaría el sensor y que permitiría su conexión o desconexión sin perturbar el normal funcionamiento de la instalación. En la figura 19.23 se muestra esta posibilidad.

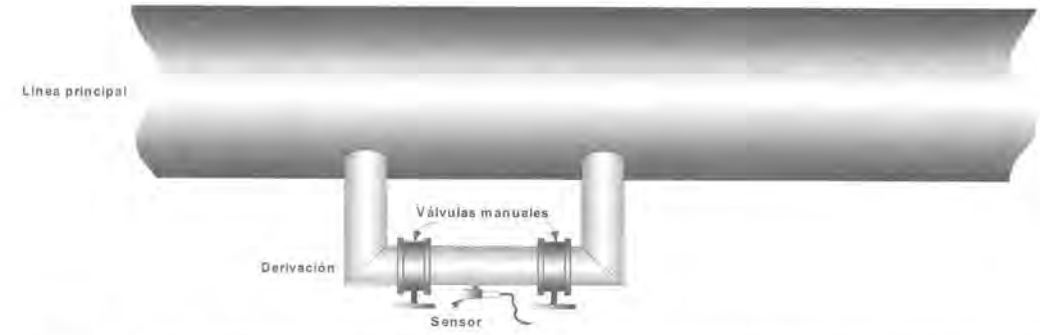


Figura 19.23. Colocación de un sensor en una derivación con válvulas para permitir su retirada sin perturbar al sistema principal.

[12] Los sensores de medida de presión suelen estar en contacto con líneas de alta presión y deben garantizar la estanqueidad de la tubería o la línea con la que están en contacto. Si el sensor es destruido o dañado, se podría ocasionar una fuga del medio que podría poner en peligro la instalación o las personas, como podría ser el caso de gases o líquidos corrosivos, explosivos o radioactivos.

El lector podría pensar que las conexiones y retiradas de los sensores no son frecuentes y que prácticamente sólo se efectúan en la puesta en marcha, cuando se produce algún tipo de avería o durante las operaciones de mantenimiento. Sin embargo, los sensores de presión suelen precisar recalibraciones frecuentes para seguir manteniendo las prestaciones iniciales. Estas recalibraciones suelen significar la retirada del sensor y la realización de una serie de pruebas que se suelen llevar a cabo en un laboratorio especializado.

19.2.7 MEDIDA DEL FLUJO

La medida del flujo que circula por una tubería o por un canal es un parámetro que tiene una gran importancia a nivel industrial, fundamentalmente en el primer caso puesto que en gran número de plantas industriales se usan u obtienen distintos tipos de fluidos (combustibles, gases, líquidos de refrigeración, productos líquidos o gaseosos, etc.) que fluyen a través de diversos conductos.

La cuantificación de lo que está ocurriendo en el interior de esos conductos se hará midiendo las condiciones del flujo que llevan lo que se consigue mediante algunos parámetros de interés:

a) **Velocidad del fluido:** magnitud expresada en m/s y que tiene interés como tal sólo en algunas ocasiones tales como canales abiertos o tuberías no llenas, casos en los que determinar otros parámetros más relevantes puede ser excesivamente complejo. La velocidad en un fluido no es un parámetro constante en cada punto de una determinada sección, sino que cambia de un punto a otro, siendo mínimo (casi nulo) en el fluido que está en contacto con las paredes. La curva de velocidad depende de un gran número de parámetros, tanto del fluido como del conducto y, por ello, se prefiere usar el *valor medio (**velocidad media**) en cada sección del conjunto. En el caso de tuberías no llenas o de* canales, la velocidad no mide la cantidad de material que fluye, sino que depende, además del área de la sección.

b) **Caudal:** esta magnitud se expresa en m^3/s e indica el volumen de fluido que circula por el conducto o canal en cada unidad de tiempo. En el caso de los líquidos o fluidos incompresibles (su volumen no se modifica con la presión) es una buena información de la cantidad de material que está circulando, mucho mejor que la anterior ya que esta sí tiene en cuenta la velocidad y la sección tanto en conductos llenos como en tuberías no llenas y canales. En el caso de los gases y fluidos compresibles, la dependencia del volumen con la presión y la temperatura hace que la medida del caudal no sea tan importante por cuanto obliga a emplear más parámetros para determinar la cantidad de material que fluye.

c) **Flujo másico:** se expresa en kg/s o en cualquier otra unidad de masa/tiempo acorde con la cantidad de fluido[13] y es la representación directa de la cantidad de materia que circula por un conducto.

Los sensores de flujo determinarán alguno de estos parámetros y, con él, representarán la cantidad de material que circula por un canal abierto o por cualquier conducción cerrada. La mayoría de las medidas de flujo se hacen de forma indirecta, a través de una variable intermedia que es controlada por el flujo debido a lo cual hay una gran cantidad de tipos diferentes de sensores de flujo basados en diversos principios de la mecánica de fluidos y de la termodinámica.

Por otra parte, las medidas de flujo suelen ser bastante complejas —y consecuentemente los sensores correspondientes— por cuanto los parámetros a tener en cuenta son muchos y de muy diversa índole. Puede afirmarse, casi sin lugar a dudas, que las medidas de flujo son las más complejas del mundo de la instrumentación en entornos industriales y que, en algunos casos, deberemos conformarnos con una aproximación que puede llegar a ser bastante grosera.

[13] Son muy habituales unidades de g/min (gramos/minuto) o, incluso mg/min cuando el fluido circula en muy pequeñas cantidades y/o muy lentamente.

A la hora de definir los criterios para seleccionar un determinado sensor para la medida de velocidad de un fluido, caudal o flujo hay que tener en cuenta algunas cuestiones que no son comunes a los procesos de selección de sensores para otras variables: para empezar, el coste no es un parámetro importante en los primeros pasos de selección y sólo debe contemplarse al final del proceso, como una comprobación: por ejemplo, la selección de un determinado sensor de caudal de bajo coste puede implicar una instalación difícil o un mantenimiento muy complejo de toda o parte de la planta con lo que el coste del sensor se llega a convertir en despreciable.

Para racionalizar el proceso de selección de sensores para medidas de flujo, vamos a fijar una serie de pasos:

1. **Tipo de información: continua** o **totalizada**. La información continua proporciona el valor del parámetro en tiempo real (velocidad, caudal o flujo másico) lo que permite conocer la situación del conducto en cada momento. En muchas ocasiones, sin embargo, resulta mucho más relevante la totalización del material que ha circulado en un determinado tiempo o durante un proceso (como en el caso de una gasolinera en que lo que importa es la cantidad total de combustible que nos van a vender). Lógicamente, no tiene sentido una medida de velocidad totalizada aunque sí que lo tiene la totalización de un caudal —en unidades de m^3 o litros— o la de un flujo másico —en unidades de kg o g—.
2. **Acceso a la información: local** o **remoto**. La medida de flujo en una planta industrial puede implicar la toma de información en lugares muy distantes entre sí o muy lejanos de un hipotético centro de control, llegándose a alcanzar varios km en algunos casos. En este tipo de situaciones o cuando es necesario leer diversas informaciones de varios puntos simultáneamente, el acceso al punto de medida no es una buena política de control y se precisará un acceso remoto a la medida; en otros casos, con plantas pequeñas o con informaciones de tipo totalizador podrá usarse un indicador local que se consultaría un determinado número de veces a lo largo de la jornada. Cuando el acceso a la información debe ser remoto, el sensor debe ser del tipo "transmisor", es decir, capaz de enviar su información a una determinada distancia usando o bien técnicas de tipo analógico o de tipo digital.
3. **Propiedades del fluido:** las propiedades físicas del fluido pueden invalidar el empleo de determinado tipo de sensores. Se pueden citar muchos ejemplos, pero podemos pensar en lo mal que trabajaría un sensor formado por una pequeña hélice para medir la velocidad en un fluido muy viscoso y que circula con lentitud. Para no caer en absurdos de este tipo —aunque no tan extremos— se deberá tener en cuenta parámetros tales como la **densidad**, la **conductividad (térmica y eléctrica)**, la **viscosidad**, la **presión de vapor** a la máxima temperatura, la **opacidad**, la **toxicidad**, la **inflamabilidad**, la **presencia de sólidos en suspensión**, lo **abrasivo** que resulte, etc. Es difícil hacer una enumeración completa de todas las propiedades que pueden determinar el tipo de sensor a emplear o impedir el uso de uno determinado o tenerlas todas en cuenta en un proceso de selección por lo que será necesario confiar en los datos de los fabricantes, en sus notas de aplicación y, sobre todo, en el sentido común.
4. **Propiedades del flujo:** las condiciones del flujo tales como la **presión**, las **temperaturas extremas** a las que pueda estar, su **reversibilidad** (posibilidad de fluir en ambos sentidos) la **presencia de burbujas** o **espuma**, el que sea **laminar o turbulento** pueden afectar a las medidas de determinados tipos de sensores o impedir el uso de otros: por ejemplo, no se puede usar un rotámetro si el flujo puede ser bidireccional o habrá que tener mucha precaución con la medida que proporciona un medidor de flujo másico en un líquido con burbujas. Cuando se habla de presión a la que está el fluido habrá que distinguir entre las condiciones normales y las extremas: por ejemplo, un dispositivo podrá trabajar correctamente en un margen de presión determinado que corresponde a las situaciones "normales" pero puede que no funcione correctamente en un caso extremo de los que no se dan casi nunca; posiblemente, en estos casos, sólo nos importará que el dispositivo soporte la situación aunque su medida se proporcione con mucho error. Un caso de este tipo son los medidores de caudal en aliviaderos de las redes de aguas residuales que trabajan proporcionando datos válidos durante todo el año pero que durante riadas o inundaciones se verán sometidos a presiones muy elevadas que deberán soportar aunque sus medidas puedan quedar fuera de alcance.

5. **Lugar de medida:** este aspecto puede ser determinante en la elección del sensor y habrá que tener en cuenta el **tamaño del conducto**, la **presencia de válvulas**, los **materiales del conducto**, la **presencia de vibraciones**, de **campos magnéticos y/o eléctricos**, el **riesgo de explosión**, la necesidad de cumplir **requisitos sanitarios**, etc. La compatibilidad general entre el sensor y el lugar de medida es algo muy importante y que puede determinar el tipo de sensor a elegir o definir las características exigibles a un tipo concreto definido por los puntos anteriores. Por ejemplo, si hemos llegado a la conclusión de que vamos a utilizar un sensor para medir la velocidad del aire en una mina de carbón, precisaremos que cumpla determinadas condiciones de seguridad (véase el último capítulo del libro) para evitar que pueda desencadenar una explosión o un incendio.

6. **Margen de medida y precisión:** aunque pueda parecer absurdo, estos parámetros que, en casi todas las ocasiones, son algo básico a la hora de seleccionar, aquí tienen una importancia menor ya que de nada sirve un sensor que reúna todas las condiciones de alcance, precisión, etc. pero que no se pueda instalar porque ponga en peligro la planta o porque, simplemente, no va poder funcionar correctamente en unas condiciones dadas. Esto significa que, por lo general, decidiremos en función de los apartados anteriores el tipo de sensor que podemos aplicar y, luego, buscaremos aquellos que cumplan las condiciones de alcance, margen de medida o precisión, aunque podría ocurrir que tuviésemos que renunciar a algún requisito en este caso y recurrir a sistemas menos precisos de lo que hubiéramos deseado.

Una vez que los anteriores parámetros hayan definido el sensor, el proceso continúa: la instalación de un sensor de flujo suele suponer una interacción directa con el fluido por lo que el proceso no es muy sencillo y, en ocasiones lleva implícito un coste elevado y un procedimiento complejo: a veces implica la modificación de una instalación existente o el rediseño de una nueva para dar cabida a un determinado medidor, para prever su recalibración o su sustitución (que implica una retirada del sensor de la instalación) con la planta en marcha. Por tanto, como decíamos al principio, el coste de un sensor debe incluir los costes asociados a todos estos procesos por lo que, una vez tomada la decisión inicial sobre el sensor, será preciso evaluarlos para saber si la decisión final va a ser o no correcta.

Los tipos de sensores para la medida de flujo (velocidad, caudal o flujo másico) se basan, como antes dijimos, en la aplicación de determinados principios de la mecánica de fluidos o de la termodinámica por lo que su explicación concreta se escapa a los objetivos del presente libro y nos limitaremos a hacer una breve enumeración de posibilidades y de ideas cualitativas de uso. Un criterio de clasificación válido puede ser el tipo de interacción que se tiene con el fluido o cuál es la magnitud intermedia que se genera; así tenemos:

a) **Sensores que miden presión diferencial generada:** se basan en el efecto que determinados obstáculos producen sobre el fluido produciendo una diferencia de presión entre un punto antes y otro después del obstáculo en función del flujo que circula (figura 19.24).

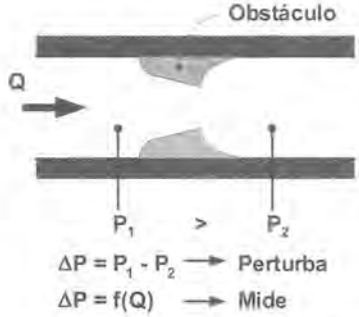


Figura 19.24. Pérdida de presión ocasionada por un obstáculo en el seno de un fluido.

La medida final que se realiza es de presión diferencial y, a partir de ella, se obtendrá la correspondiente medida de flujo a través de una expresión más o menos compleja.

Son muchos los sensores de este tipo tales como sensores de orificio, de Venturi, de cuña, de Pitot, etc. En la figura 19.25 se muestran algunos de ellos.

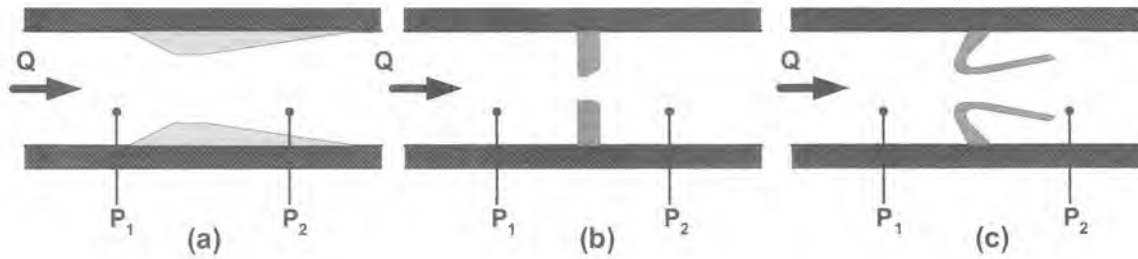


Figura 19.25. Algunos de los obstáculos habituales para la medida de caudal: (a) Venturi; (b) orificio y, (c) cuña.

El principal problema de este tipo de sensores está en la perturbación intrínseca sobre el medio en el que se está midiendo ya que la sensibilidad de la variable final que se mide —la presión diferencial— es mayor cuanto mayor sea la perturbación. El parámetro que se usa para definir esta influencia es la pérdida de presión ocasionada por la medida que suele ser baja en los sensores basados en el efecto Venturi y alta en los sensores de orificio.

b) **Sensores de sección variable.** A diferencia de los anteriores, estos sensores modifican la sección que dejan libre para el paso del fluido mediante una pieza móvil para mantener constante (más o menos) la pérdida de presión. La variable que se mide es la posición de la pieza móvil como en el caso de los rotámetros [19.21]. En la figura 19.26 se muestran algunos de estos medidores de flujo.

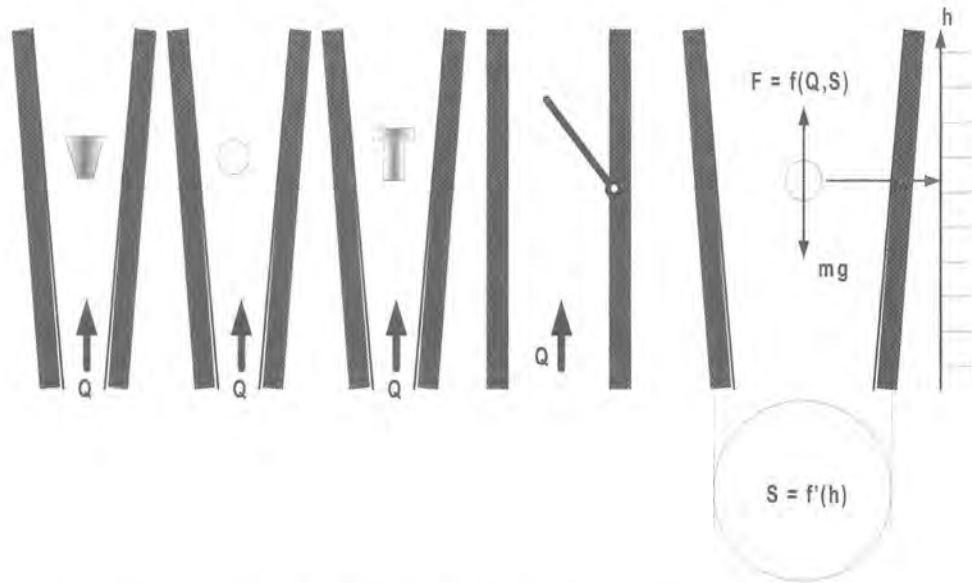


Figura 19.26. Algunos tipos de rotámetros y su principio de funcionamiento.

La variable final —desplazamiento o posición— se mide con alguno de los medios ya conocidos (potenciómetros, sistemas inductivos o capacitivos o mediante técnicas ópticas). Estos sensores deben ser situados siempre en vertical y suelen ser útiles para la medida de flujos de gases y no sirven cuando haya partículas en suspensión que puedan adherirse a la pieza móvil y causar un cambio en su tamaño, en su peso o en su superficie.

c) **Sensores mecánicos rotativos.** Emplean una pequeña turbina con álabes cuya velocidad de giro depende de la velocidad del fluido (figura 19.27).

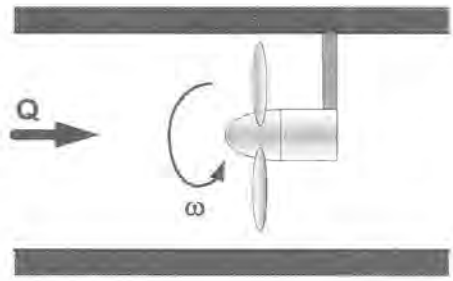


Figura 19.27. Sensor mecánico rotativo para la medida de caudal.

La velocidad de giro de la turbina, producida por la velocidad del fluido, se puede medir con cualquiera de los métodos de medida de velocidad angular comentados en este mismo capítulo. El uso de estos dispositivos es ciertamente problemático ya que, si son pequeños y perturban poco la corriente de fluido, proporcionan una información local de velocidad y que no se puede aplicar a todo la sección de fluido; si, por el contrario, son grandes, la velocidad que miden es una especie de velocidad media y tiene validez general para la sección de fluido pero ocasionan una fuerte perturbación en la línea de flujo.

Finalmente, hay que tener en cuenta que la presencia de partículas o sustancias que puedan adherirse a los álabes es muy perjudicial al modificar la función que relaciona la velocidad de giro con la velocidad del fluido; esto obliga a operaciones frecuentes de mantenimiento y limpieza lo que se ve agravado por la presencia de partes mecánicas móviles sujetas a desgaste.

d) **Sensores basados en efectos térmicos.** La transmisión de calor entre dos puntos de una corriente de fluido o la capacidad de evacuar calor de un objeto a alta temperatura pueden ser empleados para la medida de flujo: en la figura 19.28a se muestra un sensor de flujo basado en la transferencia de calor entre una fuente y un sensor térmico a través del flujo; también, en la figura 19.28b se muestra la posibilidad de medir flujo mediante la modificación de la resistencia térmica de un termistor en proceso de autocalentamiento.

Estos sensores resultan muy problemáticos cuando se pretende de ellos un uso general ya que la relación de la variable a medir (la temperatura del termistor) depende de muchos factores tales como si el flujo es laminar o turbulento, las características del fluido y de la tubería, etc. Por ello, los sensores de este tipo suelen estar calibrados para determinados fluidos y determinados flujos[14]. Otro factor que les afecta es la temperatura del fluido, la presencia de partículas en suspensión, burbujas, espuma, etc.

[14] El tipo de sensores como el que se muestra en la figura 19.28 suele emplearse en capilares con flujo laminar pero, en muchas ocasiones, se recurre a montar el sensor sobre un capilar que constituye una derivación del flujo completo. También es frecuente el emplear un gran número de capilares en paralelo para hacer que el flujo sea laminar y medir en uno de ellos.

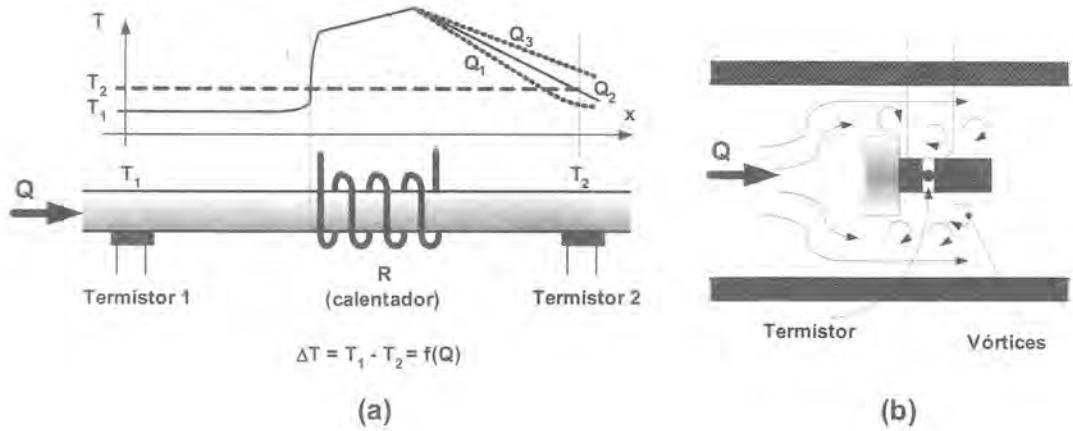


Figura 19.28. Sensores térmicos para la medida de caudal: (a) por transferencia de calor desde el calentador al sensor; (b) por modificación de la constante térmica de un termistor autocalentado.

e) **Sensores magnéticos.** Se usan en determinados fluidos que incluyen partículas cargadas sometiéndolos a un campo magnético que provoca la deflexión de las partículas y, consecuencia, la presencia de una diferencia de tensión entre extremos, en un efecto muy parecido a lo que constituye la base del efecto Hall. No es frecuente que se tengan fluidos con partículas cargadas por lo que, en muchas ocasiones, estas partículas se añaden al fluido lo que es posible siempre que el fluido no sea conductor como en el caso de los aceites.

f) **Sensores ultrasónicos.** El efecto Doppler se puede usar en la medida de la velocidad de las partículas que lleva un fluido en suspensión mediante el empleo de un emisor de ultrasonidos que trabaje a una frecuencia dada y de un receptor de ultrasonidos que recoja los ecos a frecuencia diferente (en función de la velocidad) según se muestra en la figura 19.29.

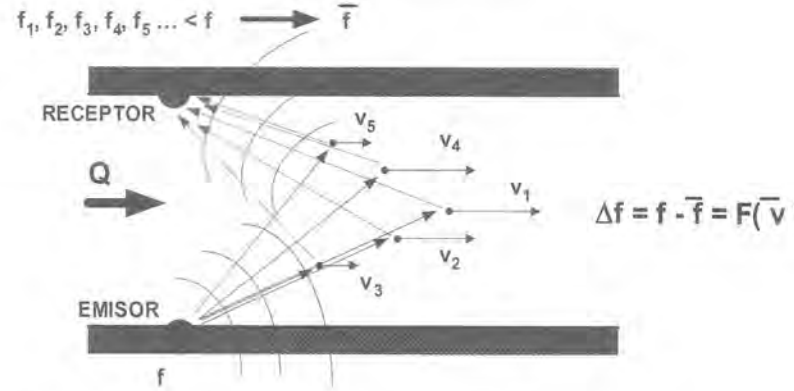


Figura 19.29. Principio de medida de la velocidad de un fluido por efecto Doppler.

Si se supone que las partículas viajan a la misma velocidad que el fluido, la medida es correcta pero esto implica una serie de limitaciones a tener en cuenta: la velocidad de las partículas puede ser muy diferente en función de su tamaño (las grandes que ocasionan ecos intensos van muy lentas o pueden precipitar) y de su posición (no todo el fluido va a la misma velocidad), lo que implica que la señal obtenida en el eco será una mezcla de diversas frecuencias procedentes de cada tipo lo

que supone la necesidad de hacer un procesado posterior para extraer una velocidad media. Lógicamente, este tipo de sensores no es útil en fluidos sin partículas en suspensión lo que limita su uso a campos muy concretos tales como caudales de aguas residuales o medidas de flujos de materiales no fluidos tales como polvos, granos, áridos, etc.

Seguro que al lector se le ocurrirá algún método más de medida de fluidos y, con casi toda probabilidad, habrá algún sensor comercial que funcione según ese método pero no vamos a alargar ociosamente una enumeración de métodos que puede ser casi inacabable.

Baste comentar, que de todos los métodos presentados, los primeros son los más "tradicionales", muy frecuentes en su uso pero con una elevada interacción con el flujo a medir por lo que tienden a no ser deseables; los últimos, tienen una baja interacción con el fluido y resultan muy útiles desde el punto de vista electrónico pero precisan una buena caracterización del flujo a medir, lo que les resta generalidad y obliga a diseños a medida que incrementan costes y dificultan la intercambiabilidad y el mantenimiento.

19.2.8 MEDIDA DEL NIVEL

El nivel del contenido de un recipiente es una variable que tiene gran interés industrial y que, por ello, suele ser muy habitual en las plantas y procesos de fabricación y de almacenaje. El sentido de la medida del nivel de material en un recipiente tiene interés por cuanto proporciona una información sobre la cantidad de material que hay en el interior del recipiente. Lógicamente, si lo que interesa es conocer el volumen de material en un recipiente o su masa (o su peso) el nivel de material no es directamente la variable de salida y, para conseguir el volumen o la masa se deberá afectar por los parámetros geométricos del recipiente y por las características de densidad del material.

Las medidas de nivel pueden corresponder a líquidos o a graneles sólidos tales como áridos, materiales pulverizados, grano, etc. La diferencia más importante entre ambas situaciones está en que la superficie libre del material puede tener formas diferentes según se muestra en la figura 19.30.

Figura 19.30. Superficie libre en un depósito: (a) líquidos estáticos; (b) graneles sólidos; (c) líquidos en agitación.

En el caso de líquidos en reposo la superficie es plana en relación con la superficie de la tierra mientras que, en el caso de graneles sólidos, puede tomar formar encopetada, con el punto más alto en un lugar dependiente de la posición de la entrada de material o de la posición de la salida del mismo. También se pueden producir superficies libres diferentes cuando los líquidos están en proceso de agitación. Este tipo de circunstancias deberán ser tenidas en cuenta para no encontrarnos con sorpresas si nuestro medidor de nivel está referido a la superficie libre del material.

En el caso de determinados líquidos y en procesos como los de fermentación o los de control de aguas residuales se pueden producir espumas y natas en la superficie libre que pueden inducir a error en determinados sensores por lo que deberá ser tenido en cuenta a la hora de elegir el dispositivo de medida.

En general, el proceso de medida de nivel tanto de líquidos como de sólidos en tanques no es tan sencillo como pudiera parecer en un principio, y esta creencia —muy extendida— puede suponer grandes errores en la medida.

Aun teniendo en cuenta las anteriores condiciones, el proceso de selección de sensores de nivel debe realizarse siguiendo una serie de pautas y respondiendo a una serie de cuestiones sobre nuestro proceso que constituyen —básicamente— los criterios de selección de sensores:

1. Se debe estudiar la **posibilidad de introducir el sensor dentro del tanque**. Si así es, cuando se retire para mantenimiento, sustitución o recalibración, debe comprobarse la posibilidad de poder extraerlo con el tanque posiblemente lleno y, tal vez, sometido a presión. Si, por el contrario, el sensor debe ser totalmente externo habrá que prever algún tipo de deriva del líquido del tanque hacia la zona en que vayamos a colocar el sensor lo que supone un incremento del coste de la instalación, más aún si se trata de una instalación sometida a presión. Colocar el sensor dentro del tanque implica necesidad de compatibilidad entre el fluido y/o los gases interiores y el sensor; esto puede ser un problema muy serio con materiales corrosivos, abrasivos o explosivos como puede ser el caso de tanques de lixiviación, tanques de combustible o silos de cereales.

2. Hay que determinar **si es posible situar el sensor inmerso en el material o en la atmósfera** que se sitúe sobre él y determinar, en cualquiera de los casos, la correspondiente compatibilidad química y física.

3. Es imprescindible determinar si el material del tanque cuyo nivel se pretende medir mantiene **constante la densidad** ya que eso afecta a la presión hidrostática y muchos sensores usan esta variable para inferir el nivel. En este sentido, cabe recordar que la densidad depende de muchos factores y que hay casos en los que se pueden encontrar variaciones importantes: la densidad varía con las condiciones batitérmicas de la atmósfera y con el tipo de líquidos. Un ejemplo claro del problema que esto supone puede ser la medida de nivel de líquido de un tanque de aceites reciclables cuyas densidades no son predecibles o un depósito de aguas residuales en un sistema de control de vertidos. Algunos sistemas de medida de nivel incluyen sensores que permiten hacer el cálculo de la densidad del fluido con lo que resultan insensibles a las variaciones en el tipo de fluido o en las condiciones batitérmicas (figura 19.31).

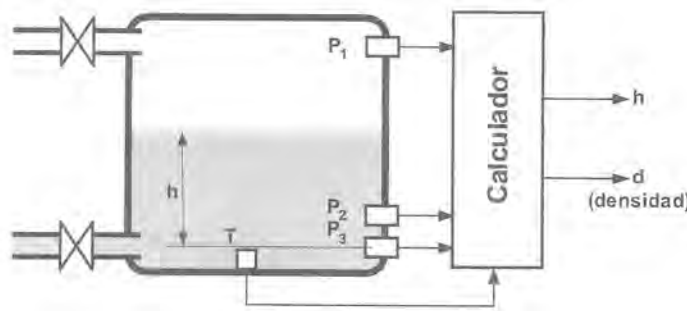


Figura 19.31. Monitorización completa de un líquido en un depósito.

4. Finalmente, se deberá tener en cuenta el **tipo de medida a realizar**: se puede desear una medida continua que proporciona una salida "proporcional" al nivel del depósito o bien una medida por

puntos que informa cuando el nivel alcanza determinados niveles. El primer caso es útil cuando el nivel forma parte de un control de procesos mientras que el segundo se da cuando lo único que se pretende es incluir alarmas de llenado en exceso, hasta determinado nivel o de depósito vacío (figura 19.32).

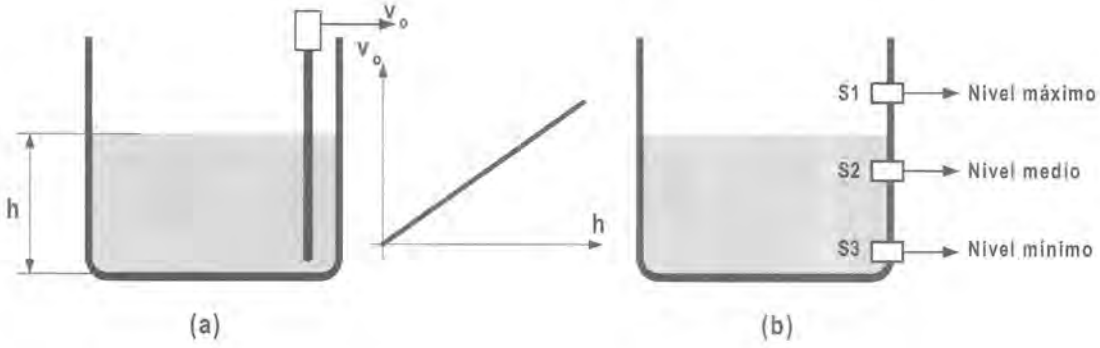


Figura 19.32. Tipos de medidas de nivel: (a) continua; (b) discreta.

En la figura 19.32a la salida del sensor resulta proporcional al nivel del depósito mientras que el caso de la figura 19.32b es el de tres sensores que se activan cuando el nivel de material llega a ellos.

A la hora de hacer un diseño en que se precise un sensor de nivel hay que tener en cuenta todas estas circunstancias, así como temperaturas de trabajo, presencia de agitadores o de otra maquinaria en el interior del tanque, los procesos de mantenimiento o, incluso, las preferencias tradicionales de algunos tipos de industria[15].

Son muchos los artilugios que se pueden pensar para hacer la medida de nivel y algunos de ellos posiblemente ya estén implementados, pero aquí vamos a centrarnos en algunos sistemas muy utilizados.

a) **Interruptores de nivel** que se activan cuando el nivel llega a ellos. Pueden estar formados por diversos tipos de sensores que van desde el empleo de flotadores hasta el uso de sensores ópticos pasando por los que miden conductividad o permitividad. Todos ellos proporcionan una salida digital o un contacto libre de tensión que informa que el nivel les ha alcanzado.

b) **Medidores de nivel** basados en cualquiera de los sensores ya estudiados y que informan sobre el nivel de material en un depósito y que se pueden construir con tecnologías muy diversas tales como los basados en la medida de la presión hidrostática, los basados en el empleo de técnicas impulso-eco de ultrasonido, los basados en láser, los capacitivos, etc. La salida de estos sensores se trata y se genera una salida de tensión o de corriente que informa al sistema del nivel real de material en el depósito.

En la tabla 19.9 se enumeran los diversos interruptores de nivel y se proporcionan algunas limitaciones y/o recomendaciones de uso.

[15] Las industrias químicas suelen preferir sensores que miden la presión hidrostática mientras que las petroleras prefieren los sistemas que miden un desplazamiento provocado por el nivel. Aunque este no sería un criterio de diseño, el conocimiento de estas limitaciones puede evitarnos algún que otro problema cuando vayamos a instalar o a diseñar una aplicación de instrumentación para alguno de estos hipotéticos clientes.

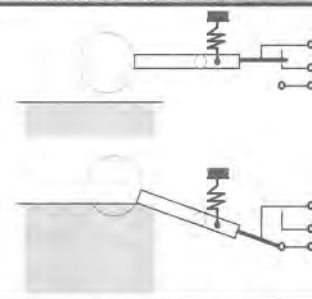
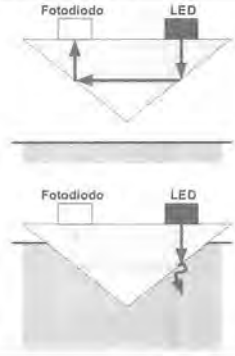
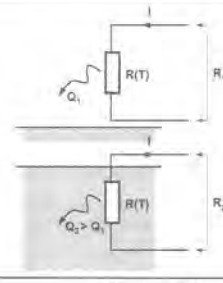
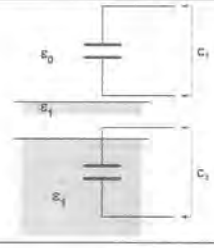
Tipo	Funcionamiento	Limitaciones y posibilidades
Flotador		– Salida por contacto libre de tensión – Vida útil limitada por desgaste – Se mancha y puede bloquearse
Óptico		– Sólo para líquidos limpios – Permite usar fibra óptica – Larga duración
Térmico		– Depende de las propiedades del fluido – Si se mancha, modifica la medida – Larga duración – Detecta espuma en superficie
Capacitivo		– Depende de las propiedades del fluido – Larga duración – Las natas pueden afectar a la medida – Diferente configuración en líquidos conductores

Tabla 19.9. Interruptores de nivel.

Los medidores de nivel se basan en los diversos efectos que el material que contiene el depósito tiene sobre una señal de excitación o en la medida de la presión hidrostática. En este último caso, la medida del nivel se hace empleando el efecto que una columna de líquido tiene sobre la presión en la parte más baja del depósito según se indica en la figura 19.33a en que se ha presentado un medidor de presión para obtener el nivel de líquido en un tanque[16].

[16] Lógicamente, el uso de los sensores de presión para la medida de nivel sólo puede hacerse para niveles de fluidos no compresibles y no en depósitos de graneles sólidos ya que la presión que medirían no se corresponde con el nivel de material.

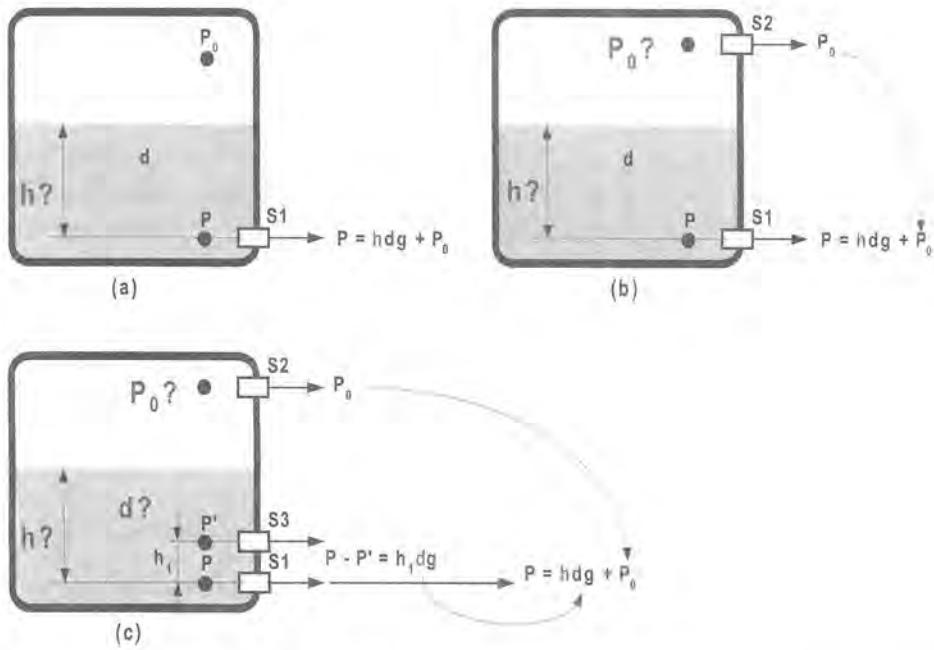


Figura 19.33. Cálculo del nivel de un líquido mediante sensores de presión: (a) si se conoce la densidad y la presión en la atmósfera sólo se precisa un sensor; (b) si no se conoce la presión en la atmósfera, se precisarán dos sensores; (c) caso general con tres sensores.

Este tipo de sensores presenta muy buenas características en cuanto a robustez y compatibilidad con diversos medios ya que los sensores de presión pueden soportar diversas condiciones de trabajo sin problema (véase, en este mismo capítulo, la parte dedicada a ellos). Sin embargo, hay varios factores que afectan a la medida y que limitan el uso de este tipo de sistemas:

a) La relación entre la presión hidrostática y el nivel del depósito es, realmente una relación entre el nivel y la presión en la atmósfera del depósito que no tiene que ser constante ni igual a la atmosférica. Por ello, debe medirse la presión en la parte superior y eliminar su efecto de la medida (figura 19.33b).

b) La presión en el punto inferior del depósito depende del nivel de columna de líquido y de la densidad. Si este último factor no se mantiene constante, bien porque el líquido no tenga unas propiedades conocidas bien porque las modifique en función de las condiciones batitérmicas del depósito. Se puede resolver el problema empleando un tercer sensor de presión separado del primero una altura conocida (figura 19.33c). Suponiendo que los sensores S1 y S3 estén sumergidos, como quiera que la altura que los separa es conocida, la diferencia de presión entre ellos permite calcular la densidad media del fluido. Con ese valor y conocida la presión diferencial entre los sensores S1 y S2, se puede obtener el nivel de líquido. También se da el caso de conocer la variación de la densidad con otro factor —generalmente la temperatura— por lo que, en este caso se emplearía un sensor de temperatura para calcular d.

Otra opción de sensores de nivel son los basados en la medida de capacidad cuyo funcionamiento aprovecha la diferencia de permitividad dieléctrica entre el medio líquido y la atmósfera según se indica en la figura 19.34 en que se muestra un condensador cilíndrico, típico en estos casos.

Figura 19.34. Sensor de nivel de tipo capacitivo.

Este tipo de sensores tiene bastantes limitaciones entre las cuales podemos citar:

a) La medida se basa en el valor de permitividad del medio líquido por lo que, si este no se mantiene constante, la medida pierde validez. Habría que recurrir a métodos de estimación de la permitividad empleando un segundo condensador de dimensiones geométricas conocidas.

b) Las mejores condiciones de medida suponen el empleo de condensadores de poco volumen (no afectan a la capacidad del depósito) y en que los radios de los electrodos son muy próximos lo que supone la aparición de efectos de capilaridad y de problemas con la espuma (modifica la medida con una permitividad dieléctrica difícil de prever) o con las natas que pueden bloquear el sensor. Por tanto, sólo se puede usar en líquidos limpios y con bajas tensiones superficiales.

c) Los líquidos deben ser no conductores y presentar una permitividad diferente de la del aire. En este caso, debería tenerse en cuenta el posible efecto de la corriente alterna en el fluido, así como la hipotética aparición de tensiones altas entre ambos electrodos lo que puede ocasionar arcos con la peligrosidad que eso implica en determinadas condiciones.

d) Para líquidos conductores se puede modificar el sistema de medida haciendo que uno de los electrodos sea el propio líquido lo que puede plantear problemas al hacer circular corriente por él.

e) La presencia de agitadores debe ser tenida en cuenta a la hora de buscar una situación idónea del sensor en el tanque ya que puede interferir con ellos.

f) La forma de los depósitos o tanques puede ser un inconveniente para la medida y habrá que tenerlo en cuenta para calcular el nivel a partir de la capacidad del sensor.

Este tipo de sensores no es válido para la medida de graneles sólidos aunque tengan una permitividad dieléctrica del material diferente de la de la atmósfera debido a que no es posible garantizar que el nivel dentro del condensador de medida y fuera del mismo sean iguales; sin embargo, sí que se utilizan interruptores de nivel capacitivos en depósitos de graneles sólidos.

Otra forma de medir niveles es mediante técnicas de impulso-eco de ultrasonidos según se comentó en la lección correspondiente a los sensores piezoeléctricos, obteniendo funcionamientos bastante buenos incluso en casos de atmósferas poco controladas o sucias. Este tipo de sistemas tiene algunos inconvenientes cuando se usan en depósitos de sólidos en que la superficie libre no es plana o en depósitos con agitadores debido a la posible presencia de falsos ecos procedentes de la agitación.

Técnicas similares basadas en el empleo de microondas o radiación luminosa permiten obtener sensores de nivel bastante precisos aunque imponen algunas condiciones de limpieza bastante severas (en el caso de la luz) [19.21].

19.3 Medida de magnitudes eléctricas

Los dispositivos y tecnologías más comunes para la medida de magnitudes eléctricas (tensión, corriente y potencia) son principalmente: resistencias *shunt*, transformadores de intensidad, sensores de efecto Hall, sensores magnetorresistivos y sensores de fibra óptica. Todos ellos, excepto los transformadores de medida, han sido estudiados en capítulos anteriores, por lo que el objetivo de este capítulo es el de presentar su aplicabilidad a la medida de magnitudes eléctricas, destacando las virtudes y limitaciones que cada uno de ellos presentan.

19.3.1 RESISTENCIA *SHUNT*[17]

La forma más simple de medir la corriente eléctrica es colocar una resistencia en serie con la carga y medir la caída de tensión en ella. Este método permite realizar, con bajo coste, medidas tanto en continua como en alterna. A fin de minimizar la disipación de potencia en dicha resistencia, su valor ha de ser el menor posible, lo cual conlleva la necesidad de una amplificación, generalmente de tipo diferencial. Un inconveniente importante es la falta de aislamiento galvánico entre el circuito de potencia y el de medida, lo cual obliga al uso de amplificadores de aislamiento, encareciéndose el diseño.

La figura 19.35 muestra una sencilla alternativa para la medida de pequeñas corrientes, empleando amplificadores operacionales. Como se sabe, la tensión de salida de este circuito está dada por la expresión: $v_0 = I\,R_S(1+R_2/R_1)$. Valores bajos de la resistencia R_S proporcionan una mayor exactitud y estabilidad térmica y menor constante de tiempo (mayor rapidez en la medida).

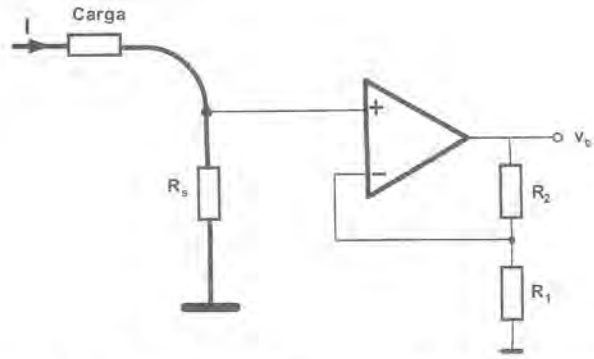


Figura 19.35. Configuración básica para la medida de corriente.

Existen soluciones integradas que permiten monitorizar la corriente por una resistencia *shunt*. Así, la figura 19.36 muestra la arquitectura interna del INA139 de Texas Instruments [19.28]. La corriente a medir

[17] El nombre de resistencia *shunt* es un tanto equívoco, puesto que la resistencia de medida se coloca en serie con la carga y no en paralelo.

circula por la resistencia *shunt* R_S. La función de transferencia del INA139 es: $I_O = g_m(V_{IN}^+ - V_{IN}^-)$, donde la transconductancia, g_m, está dada por $g_m = 1000\ \mu A/V$. Como $V_{IN}^+ - V_{IN}^- = I_S \cdot R_S$ y la tensión de salida $v_o = I_O \cdot R_L$, la función de transferencia completa de este circuito es:

$$v_o = [1000\ \mu A/V \cdot I_S \cdot R_S] \cdot R_L \qquad (19.1)$$

La selección de R_S depende de la aplicación, resultando ser un compromiso entre la exactitud deseada y la caída de tensión máxima permitida en la línea de medida. Valores altos de R_S proporcionan mejor precisión para corrientes más bajas, minimizando los efectos del *offset*, mientras que valores bajos de R_S minimizan la caída de tensión en la línea de alimentación. Se recomienda que la tensión de entrada diferencial esté entre 50 y 100 mV (máximo 0,5 V). El valor de R_L se elige para obtener la tensión de salida de fondo de escala deseada. La impedancia de salida del INA139 es muy alta, lo que permite usar valores de R_L hasta 100 kΩ, con excelente exactitud.

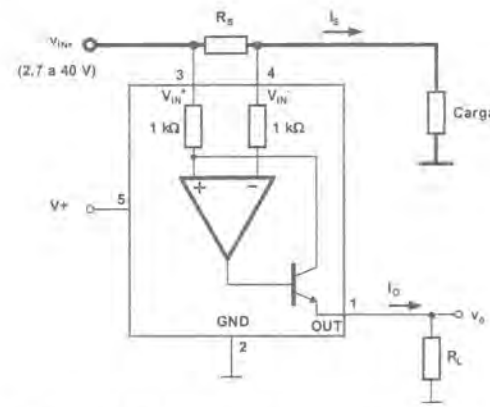


Figura 19.36. Esquema interno del INA139.

EJERCICIO 19.1

A partir del circuito INA139 de la figura 19.36 determine los valores adecuados de la resistencia shunt R_S, y de la resistencia de carga R_L si se quiere que, cuando la corriente a medir sea de 1 A, la tensión de salida de fondo de escala sea 100 mV.

SOLUCIÓN:

El valor de R_S se elige de forma que la tensión diferencial de entrada $V_{in}^+ - V_{in}^-$ sea 100 mV, con lo que $R_S = 0{,}1/1{,}0 = 0{,}1\ \Omega$. Es aconsejable utilizar una resistencia de película metálica.

El valor de R_L se selecciona para obtener una tensión de salida de fondo de escala de 0,1 V, por lo tanto, aplicando la expresión (19.1), resulta $R_L = 0{,}1\ V/(1000\ \mu A/V \cdot 1A \cdot 0{,}1\ V) = 1\ k\Omega$.

Otro circuito integrado muy popular es el ZXCT1009 de la firma Zetex [19.32]. Es un circuito de bajo coste, con una arquitectura interna igual a la del INA139, exactitud típica del 1% y ancho de banda de 300 kHz. Su función de transferencia es: $v_{out} = 0{,}01 \times R_S I_S \times R_{out}$. En aplicaciones de bajo coste, la resistencia *shunt* R_S, se puede implementar como una pista de cobre, cuando se realiza el PCB. La figura 19.37a muestra el circuito de aplicación y la figura 19.37b el trazo de cobre usado para implementar R_S. En la hoja de

datos del fabricante pueden consultarse otras aplicaciones de este circuito como monitorizar corrientes bidireccionales o protección contra cortocircuitos.

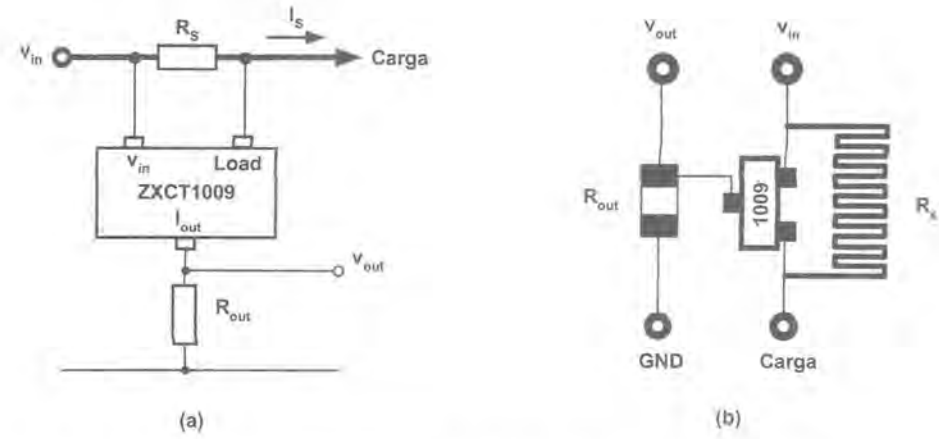


Figura 19.37. Uso del ZXCT1009 para medir corriente: (a) circuito eléctrico; (b) implementación de R_S en las pistas de cobre del PCB.

19.3.2 TRANSFORMADOR DE INTENSIDAD[18]

Un transformador de intensidad permite detectar corrientes alternas, con aislamiento galvánico. Está constituido por un devanado primario (generalmente de una espira) por el que circula la corriente a medir y un devanado secundario al que se conecta una carga que puede ser una resistencia de medida o un aparato de medida (figura 19.38a). Las principales limitaciones que presenta son que no permite medir señales continuas y que tiene un precio relativamente alto. Por el contrario su dependencia con la temperatura es prácticamente despreciable. En un transformador de intensidad ideal, al igual que en uno de potencia, se cumple que los amperios vuelta de los devanados primario y secundario son iguales, es decir: $N_P I_P = N_S I_S$, de donde se tiene que la corriente por el secundario será:

$$I_S = I_P \cdot \frac{N_P}{N_S} \tag{19.2}$$

Es habitual que $N_P = 1$, con lo que la corriente primaria se verá reducida en el secundario por el número de vueltas del devanado secundario, N_S. La expresión (19.2) será tanto más válida cuanto menor sea la corriente de vacío o de excitación, para lo cual se emplean materiales de alta permeabilidad magnética. Se define la **relación de transformación nominal** k_n, como el cociente entre los valores eficaces de las intensidades nominales primaria y secundaria:

$$k_n = \frac{I_{Pn}}{I_{Sn}} \tag{19.3}$$

[18] Para la medida de tensiones de alto valor puede emplearse un **transformador de tensión**. Su construcción es similar a la de un transformador de potencia funcionando en vacío ya que está pensado para trabajar conectado a un aparato de medida con una impedancia de entrada de valor muy elevado.

La corriente de excitación da lugar a dos tipos de errores: el **error de relación** ε, definido como el error que el transformador introduce en la medida de una corriente, como consecuencia de que la relación de transformación real no es igual a la relación de transformación nominal.

$$\varepsilon\ (\%) = \frac{k_n I_S - I_P}{I_P} \times 100 \qquad (19.4)$$

y el **error de fase** δ, definido como la diferencia de fase entre los vectores de las corrientes primaria y secundaria (figura 19.38b).

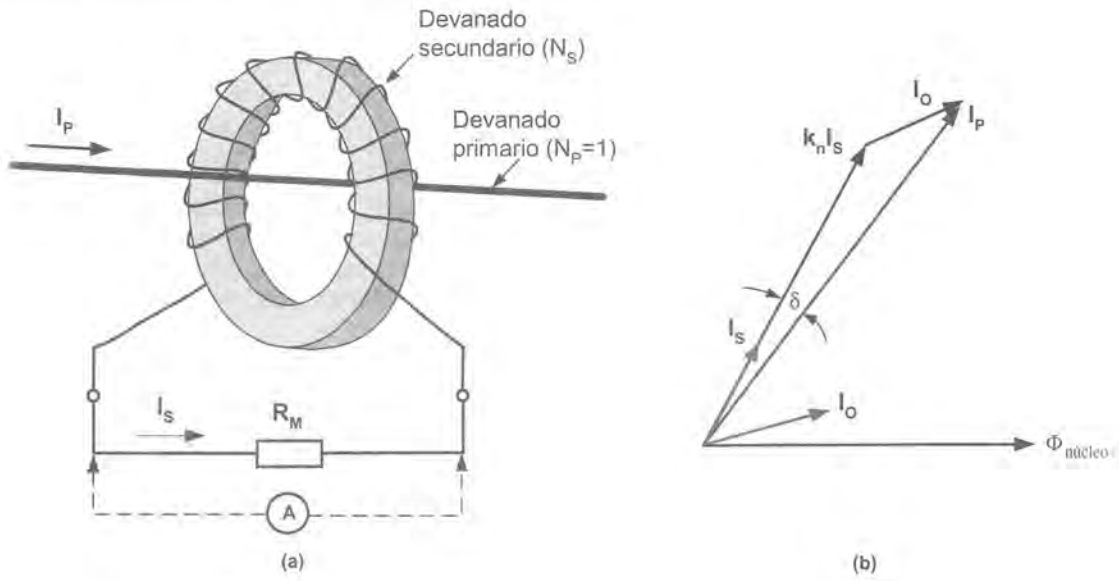


Figura 19.38. Transformador de corriente: (a) construcción y (b) diagrama vectorial.

Las características más importantes de cara a seleccionar un transformador de intensidad son principalmente:

- **Corriente nominal primaria:** se elige en función de la corriente a medir. Los valores normalizados para transformadores dedicados a alimentar aparatos de medida son: 10 – 12,5 – 15 – 20 – 25 – 30 – 40 – 50 – 60 – 75 A.
- **Corriente nominal secundaria:** las salidas normalizadas son 1 y 5 A. También es posible tener una salida 4-20 mA para control de procesos.
- **Clase de precisión:** se define como el error de relación máximo, para la intensidad primaria nominal y la carga de precisión. La clases de precisión normales son: 0,1 – 0,2 – 0,5 – 1 – 3.
- **Carga de precisión:** es el valor de la carga que garantiza las especificaciones de precisión anteriores. Se suele expresar por su valor en ohmios y su factor de potencia o por la potencia aparente (VA) que consume al ser atravesada por la corriente secundaria nominal.

Nunca debe dejarse el secundario en circuito abierto ya que en tal caso toda la corriente primaria pasa a ser corriente de excitación, produciendo la saturación magnética del núcleo y dando lugar a la aparición de una tensión en los terminales del secundario, que puede ocasionar fallos en el aislamiento del transformador,

además de ser un peligro para el usuario. Por tanto, si por algún motivo hay que desconectar la carga, se tiene que "puentear" el secundario.

La figura 19.39 muestra el conexionado de un transformador de intensidad para la medida de la potencia eléctrica de un sistema. Hay que fijarse en el sentido del conexionado de los terminales del transformador, ya que si cambiásemos el terminal S1 por el S2, la corriente en el secundario estaría desfasada 180° y la lectura del vatímetro sería incorrecta.

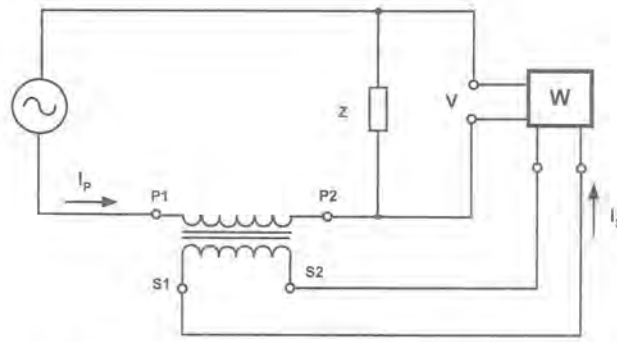


Figura 19.39. Medida de la potencia eléctrica usando un transformador de intensidad.

19.3.3 SENSORES DE EFECTO HALL

En el capítulo 17 se estudiaron el fundamento y aplicaciones de los sensores de efecto Hall. Estos dispositivos permiten monitorizar tanto señales continuas como alternas proporcionando aislamiento galvánico entre el circuito de medida y el de potencia.

La medida de corriente (o de tensión) puede realizarse en bucle abierto o en bucle cerrado. A partir de *los datos ofrecidos por la tabla [19.16], se puede afirmar que la medida en bucle cerrado aporta una mejoría* en todas las características técnicas: exactitud, linealidad, derivas. *Esto se debe a que integra un circuito de* compensación que mantiene nulo el flujo magnético en el núcleo, con lo que se elimina la influencia de las no linealidades del núcleo y se reducen los efectos de la temperatura en el elemento Hall (véase apartado 17.5).

Característica	Medida de corriente en bucle abierto	Medida de corriente en bucle cerrado
Margen de medida	0-18000 A	0-15000 A
Ancho de banda	0-25 kHz	0-200 kHz
Exactitud a 25°C	±1%	±0,5%
Linealidad	±0,5%	±0,1%
Tiempo de respuesta	< 3-7 μs	<1 μs
Margen de temperatura	–25 a +70°C	–40 a +85°C

Tabla 19.10. Comparación de sensores Hall de corriente.

La figura 19.40 muestra tres versiones comerciales de sensores Hall de corriente: para montaje en PCB, núcleo sólido y núcleo partido. Este último se utiliza para aplicaciones de sustitución, donde no sea posible

desconectar el cable del primario. Los dispositivos para montaje en PCB pueden llevar un orificio para el conductor por el que circula la corriente o bien la corriente se lleva a los terminales de entrada. La mayoría de las versiones admiten múltiples vueltas del primario, al objeto de poder incrementar la sensibilidad de la medida.

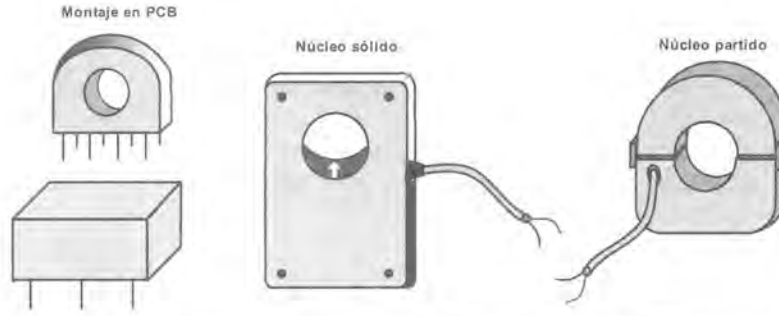


Figura 19.40. Formas constructivas de sensores Hall de corriente.

La salida puede ser en corriente (figura 19.41a) o en tensión (figura 19.41b). En el caso de salida en corriente, esta se mide sobre una resistencia exterior R_M cuyos valores máximos y mínimos suelen venir dados por el fabricante. Las corrientes de salida típicas son 25, 50 y 100 mA para entradas nominales de 25, 50 y 100 A, respectivamente. También son frecuentes las versiones con salidas 4-20 mA, de forma tal que cuando la corriente de entrada sea nula, la salida tome el valor de 4 mA y cuando esté pasando la corriente máxima, la salida sea de 20 mA. La salida en tensión puede ser unipolar 0-10 V, idóneas para su uso con amplificadores de instrumentación y medidores de panel, o bien bipolar −10 V – 0 – +10 V, apropiada para aplicaciones donde sea necesario el control de corrientes alternas. En la referencia [19.18] se describe la aplicación de un sensor de efecto Hall para la medida de una corriente máxima de 220 A.

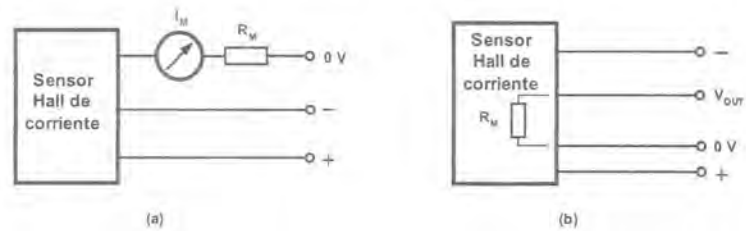


Figura 19.41. Conexiones típicas de un sensor Hall de corriente: (a) salida en corriente y (b) salida en tensión.

EJERCICIO 19.2

Un dispositivo Hall de corriente en bucle cerrado tiene las siguientes características:

- *Corriente nominal primaria (rms): $I_{PN} = 50$ A*
- *Margen de medida de la corriente primaria (rms): $I_P = 0 \pm 70$ A*
- *Resistencia de medida: $R_{M\,min} = 10\ \Omega$, $R_{M\,máx} = 100\ \Omega$ ($T_A = 70°C$, $V_C = \pm 12$ V)*
- *Relación de conversión: $K_N = 1:1000$*
- *Corriente nominal secundaria: $I_{SN} = 50$ mA*
- *Resistencia de la bobina secundaria: $R_S = 80\ \Omega$*

Determine:

a) El valor de la resistencia de medida necesaria para obtener una tensión de medida de 3,5 V para la corriente nominal primaria.

b) ¿Cuál es la tensión máxima de medida que se puede obtener con este dispositivo?

SOLUCIÓN:

a) Como la relación de conversión es 1:1000, la corriente nominal secundaria correspondiente a la corriente nominal primaria es de 50 mA. La resistencia de medida necesaria es de:

$$R_M = \frac{V_M}{I_S} = \frac{3{,}5\ V}{50\ mA} = 70\ \Omega$$

b) La tensión máxima se obtiene para la corriente máxima de primario de 70 A (en el secundario 70 mA) y resistencia de medida de 100 Ω. Es decir:

$$V_M = R_M I_S = 100\ \Omega \cdot 70\ mA = 7\ V$$

Medida de tensión. La medida de tensión mediante sensores Hall está basada en el mismo principio que la medida de corriente, de hecho, el sensor Hall de tensión está constituido por un sensor de corriente. La principal diferencia está en el circuito primario, en el que se ha aumentado el número de espiras, de forma que la corriente primaria sea baja, con un consumo mínimo en el circuito de medida.

Para medir una tensión es suficiente con llevar al sensor una corriente primaria equivalente mediante una resistencia en serie con el devanado primario. La figura 19.42 muestra dicha conexión para el caso del sensor Hall de tensión [19.16]. La resistencia R_1 puede estar integrada en el propio dispositivo.

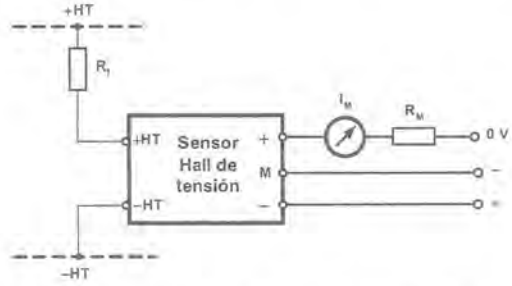


Figura 19.42. Conexionado típico de un sensor Hall de tensión.

Medida de potencia. La potencia eléctrica disipada por una carga en un circuito de continua se obtiene como el producto de la tensión y de la corriente, es decir: $P_L = V_L \cdot I_L$. El problema, en continua, se reduce por tanto, a medir el valor de la tensión y de la corriente y a realizar su producto.

En el caso señales alternas, la potencia eléctrica disipada está dada por el producto de los valores instantáneos. En circuitos alimentados por señales periódicas interesa medir la potencia activa P, definida como el valor medio de la potencia instantánea:

$$P = \frac{1}{T}\int_0^T p(t)dt \qquad (19.5)$$

El cálculo de la expresión 19.5 puede hacerse mediante el esquema de la figura 19.43, en el que se acondicionan las señales de corriente y tensión, se multiplican sus valores y se calcula su valor medio mediante un integrador.

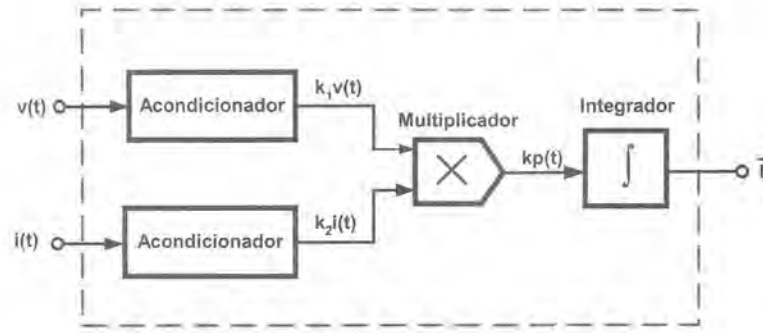


Figura 19.43. Diagrama de bloques de una medida de potencia activa.

Es posible realizar un multiplicador basado en el efecto Hall. La figura 19.44 muestra el principio de funcionamiento de esta técnica. El campo magnético B, es proporcional a la corriente que circula por la carga i_L, es decir: $B(t) = k_1\, i_L(t)$. La corriente de control del sensor Hall i_C, es proporcional a la tensión de línea: $i_C(t) = k_2\, v_L(t)$. La tensión Hall está dada por: $v_H(t) = k_H \cdot i_C(t) \cdot B(t)$, donde k_H es la constante de Hall. Por tanto, el valor instantáneo de la tensión Hall será: $v_H(t) = k_H \cdot i_C(t) \cdot B(t) = k \cdot v_L(t) \cdot i_L(t)$, es decir, la tensión Hall es proporcional a la potencia en la carga. Esta señal analógica puede ser procesada para obtener la potencia media o activa.

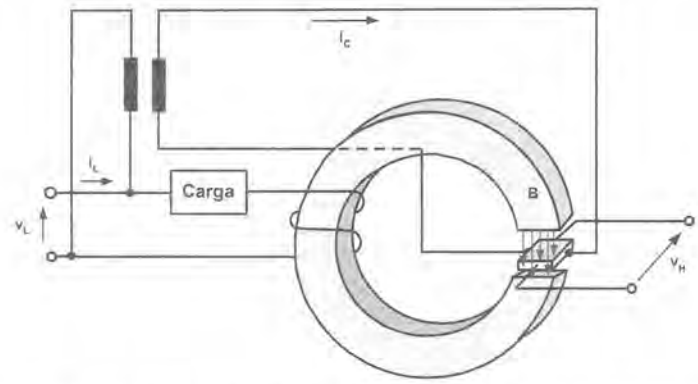


Figura 19.44. Medida de potencia eléctrica basada en el efecto Hall.

19.3.4 SENSORES MAGNETORRESISTIVOS

Los sensores magnetorresistivos (MR) fueron estudiados en el capítulo 9. Si bien tanto los sensores de efecto Hall como los magnetorresistivos requieren la presencia de un campo magnético, hay que recordar que [19.11]:

- Los MR son más sensibles (unas 200 veces) que los sensores de efecto Hall, si bien su linealidad es peor. Esta linealidad se mejora sustancialmente mediante una técnica de compensación similar a la empleada por los sensores Hall de corriente en bucle cerrado: se integra en la misma estructura un conductor de compensación que al circular por él la corriente crea un campo magnético que compensa el creado por la corriente a medir (figura 3 de la referencia [19.17]).

- Los MR tienen un ancho de banda mayor: entre 1-15 MHz frente a unos 200 kHz de sensores Hall.
- En los sensores Hall, el eje sensible es perpendicular al plano del chip, en cambio, en los MR, el eje sensible es paralelo al plano del chip.
- Ambas tecnologías son compatibles con las tecnologías de fabricación de circuitos integrados.

La disposición interna que suelen presentar los sensores MR es la de un puente de Wheatstone con dos o cuatro MR. En cada rama del puente las MR tienen sensibilidad opuesta a la acción del campo magnético, lo que permite reducir su dependencia con la temperatura, así como la influencia de campos magnéticos externos.

La corriente eléctrica (continua o alterna) puede ser medida con un sensor MR, mediante la detección del campo magnético que dicha corriente crea. La figura 19.45 muestra un ejemplo de sensor magnetorresistivo **AMR** de la firma Zetex [19.33], en el que la corriente a medir I_M, circula por un conductor en el interior del chip. La salida del puente se amplifica y filtra para dar una tensión de salida proporcional a la corriente a medir. Este fabricante dispone de modelos con corrientes de fondo de escala de 5, 10 y 20 A y un ancho de banda de 100 kHz.

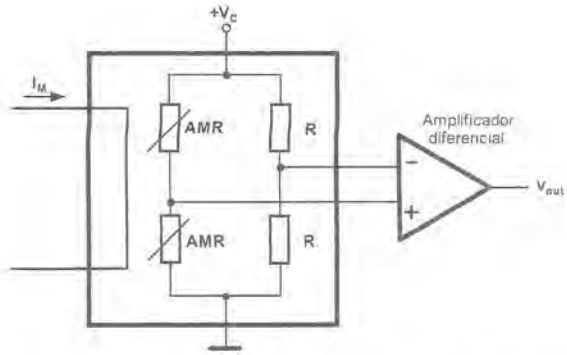


Figura 19.45. Estructura interna de un sensor de corriente MR.

Los sensores magnetorresistivos del tipo **GMR** permiten obtener variaciones de resistencia mayores (entre un 4-20%), obteniéndose un nivel de salida mayor que con los AMR. La configuración típica de un sensor GMR es la de un puente de Wheatstone con cuatro resistencias GMR, de las cuales dos son sensibles al campo y las otras dos están apantalladas. La figura 19.46 muestra un ejemplo de sensor GMR, de la firma *NVE Corporation* [19.20]. La tensión de salida máxima es del orden de 0,3 V por lo que suele ser necesaria una amplificación posterior.

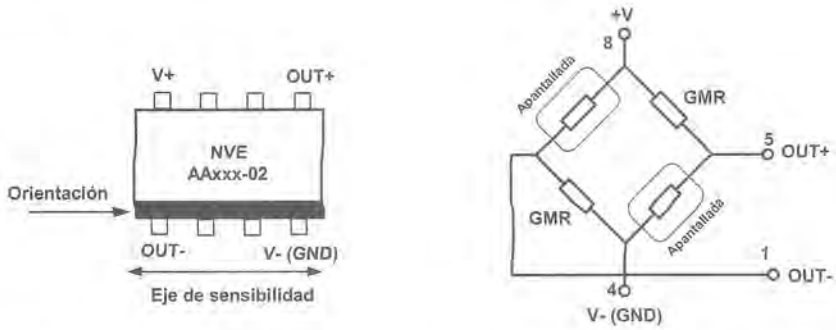


Figura 19.46. Patillaje y diagrama de bloques del sensor GMR de la serie AA002 de NVE.

La figura 19.47a muestra el empleo de un sensor GMR para medir la corriente que circula por un conductor situado a una distancia, D, de la superficie donde está el sensor. En la figura 19.47b se trata de medir la corriente que circula por una pista de circuito impreso. En la referencia [19.20] puede encontrarse una hoja de cálculo para obtener el valor del campo magnético y la tensión de salida del sensor en función del valor de la corriente, las distancias representadas y la tensión de alimentación del sensor.

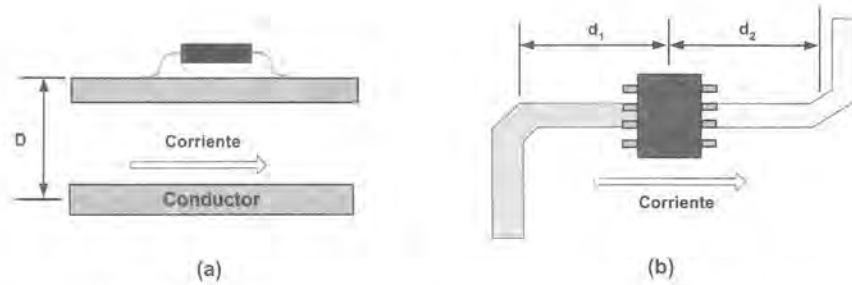


Figura 19.47. Medida de la corriente eléctrica mediante un sensor GMR.

Las aplicaciones típicas de los sensores de corriente MR se encuentran en: automoción (medida de consumo de corriente en el circuito de alumbrado de los coches, sistemas de frenado ABS, etc.), sistemas electrónicos de potencia (inversores, fuentes de alimentación conmutadas, SAIs, etc.), gestión de baterías (corrientes de carga y descarga de baterías), vehículos eléctricos, tecnología solar, robótica, etc.

19.3.5 SENSORES DE FIBRA ÓPTICA

Los sensores de fibra óptica fueron presentados en el capítulo 18. Su aplicación para monitorizar señales eléctricas es una consecuencia de las características específicas que presentan estos sensores: no son conductores, tienen poco peso (lo que facilita su aislamiento y montaje), no presentan histéresis y poseen una buena respuesta en frecuencia. Son particularmente interesantes en aplicaciones de alta tensión o en presencia de interferencias electromagnéticas y donde la velocidad y estabilidad sean características importantes [19.6].

La figura 19.48 muestra un sensor de corriente intrínseco basado en la modulación de la intensidad óptica. La corriente I que circula por el conductor, induce un campo magnético, el cual, en virtud del **efecto Faraday**[19], rota el plano de polarización de la luz, modulando la intensidad de la luz transmitida de forma proporcional al campo magnético y, por tanto, a la corriente a medir.

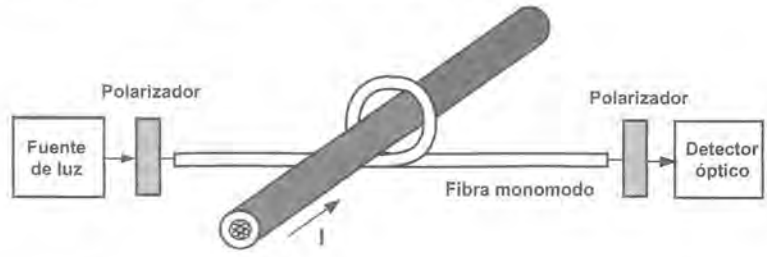


Figura 19.48. Sensor óptico de corriente intrínseco.

[19] El efecto Faraday consiste en que un campo magnético, aplicado a ciertos materiales, da lugar a una actividad óptica si dicho campo magnético tiene una componente en la dirección de propagación de la radiación en el material. Este efecto aparece en muchos materiales, pero sólo en los diamagnéticos es independiente de la temperatura y puede utilizarse para la medida de corrientes.

El ángulo de rotación θ del plano de polarización, se define como:

$$\theta = V \int B \cdot dl \tag{19.6}$$

donde V es un parámetro del material conocido como constante de *Verdet*, B es el vector densidad del flujo magnético y dl es el vector diferencial a lo largo de la dirección de propagación.

Para un conductor en lazo cerrado por el que circula una corriente I, la Ley de *Ampère*, establece que:

$$\oint H \cdot dl = I \tag{19.7}$$

por lo que el ángulo de rotación, está dado por:

$$\theta = \mu VNI \tag{19.8}$$

donde μ es la permeabilidad magnética del material y N el número de espiras de fibra óptica. Por tanto, midiendo el ángulo de rotación y conociendo la constante de *Verdet*, es posible conocer el valor de la corriente que circula por el conductor. En principio, el sensor es insensible a los campos magnéticos parásitos debidos a conductores cercanos. Basado en el efecto Faraday, la firma NxtPhase [19.6] desarrolló, para aplicaciones de alta tensión, el sensor de corriente de fibra óptica de la figura 19.49. El resultado ha sido un sensor óptico más barato que un transformador de corriente, sin histéresis y con mayor margen dinámico.

Es posible medir un campo magnético (y, por tanto, una corriente eléctrica) usando una fibra óptica con un recubrimiento de un material magnetoestrictivo. Estos materiales presentan pequeños cambios en sus dimensiones físicas (del orden de algunas ppms) en presencia de un campo magnético. El cambio en la longitud de la fibra asociado a un cambio en el campo magnético da lugar a un desplazamiento de fase $\Delta\phi$, que puede ser detectado mediante técnicas interferométricas (véase capítulo 18).

La figura 19.50 muestra un ejemplo de esta técnica [19.7], consistente en una sección de fibra óptica situada en el centro de un solenoide por el que circula la corriente que está siendo medida. Mediante esta técnica se puede llegar a medir una corriente eléctrica mínima de 30 nA/m en un margen de frecuencias de 100 a 500 Hz. La respuesta frecuencial depende fuertemente del material, así como de la geometría. Por encima de, aproximadamente 1 kHz, se observa una disminución de la sensibilidad.

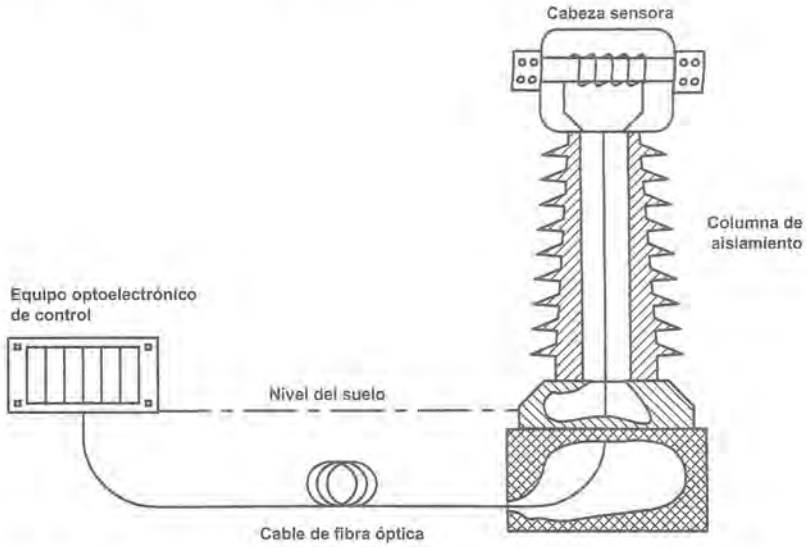


Figura 19.49. Instalación típica de un sensor de fibra óptica para la medida de corriente.

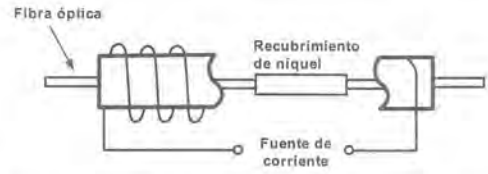


Figura 19.50. Sensor de fibra óptica para la medida de corriente eléctrica.

A modo de resumen, la tabla 19.11 compara los principales aspectos de las tecnologías consideradas en este capítulo para monitorizar magnitudes eléctricas [19.15]. Se trata de una información de tipo general, que habrá que matizar en función del tipo de aplicación concreta en la que se vaya a utilizar dicha tecnología.

	Resistencia *shunt*	Transformador de corriente	Sensor de efecto Hall	Sensor magnetorresistivo	Sensor de fibra óptica
Coste	Muy bajo	Medio	Alto	Alto	Medio
Linealidad	Muy buena	Media	Media	Media	Buena
Capacidad de medir corrientes altas	Muy pobre	Buena	Buena	Baja	Muy buena
Consumo	Alto	Bajo	Medio	Bajo	Medio
Variación de la salida con la temperatura	Media	Baja	Alta	Media	Baja
Offset	Sí	No	Sí	Sí	No
Histéresis y saturación	No	Sí	Sí	Sí	No

Tabla 19.11. Comparación de tecnologías para la medida de corriente eléctrica.

19.4 Medida de magnitudes ópticas

En esta sección se estudian los parámetros más importantes de los sensores ópticos. El estudio de dispositivos ópticos de formación de imágenes (espejos y lentes) y de algunos instrumentos de medida (espectrómetros, radiómetros, etc.) pueden verse en el Anexo 8 del CD adjunto.

19.4.1 SISTEMAS DE MEDIDA DE RADIACIÓN ÓPTICA

La radiación óptica comprende un margen de longitudes de onda del espectro electromagnético que se extiende desde 100 hasta 10.000 nm (desde el ultravioleta al infrarrojo). Las medidas que generalmente se realizan están relacionadas con medidas de la longitud de onda de la radiación (espectrómetros), medidas de color (espectrofotómetros o colorímetros) y medidas de densidad de flujo (fotómetros o radiómetros).

Espectrómetros

Las medidas del espectro de una fuente de luz pueden realizarse con diversos tipos de instrumentos. En las medidas denominadas ratiométricas de banda ancha (*broad-band*) se utilizan detectores corregidos, es decir, se corrige la respuesta de un sensor por medio de filtros (véase apéndice 8 en el CD) para obtener el componente espectral deseado. Este tipo de medidas son, en general, bastante fáciles de realizar y los sistemas que

resultan no son muy caros. La principal desventaja reside en la dificultad para diseñar un filtro que se acomode de forma exacta a la respuesta buscada. Otro problema asociado con estas medidas está relacionado con la resolución espectral que proporcionan, que suele ser bastante baja. Para medidas más precisas del espectro se utilizan otros sistemas que constan básicamente de un sistema óptico que recibe y limita la radiación a medir, un monocromador (véase apéndice 8), un sensor óptico y el sistema de acondicionamiento y de visualización correspondiente. Un sistema de este tipo es el denominado **espectro-radiómetro** (figura 19.51) utilizado para la medida de la distribución de potencia espectral (potencia radiada por longitud de onda) de una fuente de luz. La esfera que aparece en la figura 19.51 se denomina esfera integradora[20]. Es una esfera hueca, cubierta por un material difusor con gran capacidad reflectante. En este caso, la fuente se sitúa externa a la esfera y esta se utiliza para homogenizar la fuente de radiación y evitar puntos intensos de luz (*hot-spots*) que perturben la medida. El monocromador permite seleccionar una determinada longitud de onda que se obtiene por su rejilla de salida y se aplica al elemento sensor. La respuesta espectral en estos sistemas de medida depende de la naturaleza del detector. Algunos utilizan PMTs por su alta sensibilidad y porque su respuesta se extiende aproximadamente entre 125 y 1100 nm. También se utilizan fotodiodos de silicio que cubren el margen 200-1200 nm. Para medidas en el infrarrojo pueden utilizarse diodos de germanio (1000-4000 nm) o de arseniuro de indio (1000-3600 nm). En muchas ocasiones los espectro-radiómetros suelen utilizarse en una configuración denominada **espectro-gráfica**, donde la rendija de salida se reemplaza por una película fotográfica. Al eliminar la rendija que limita la apertura, la radiación se dispersa y se proyecta sobre la película obteniendo una fotografía de los componentes espectrales. En los instrumentos modernos, sin embargo, y tal como se indica en la figura 19.51, la película se sustituye por un **conjunto de fotodiodos** (*photodiode array*). Cada fotodiodo detecta la intensidad de la radiación incidente en una banda estrecha de longitudes de onda.

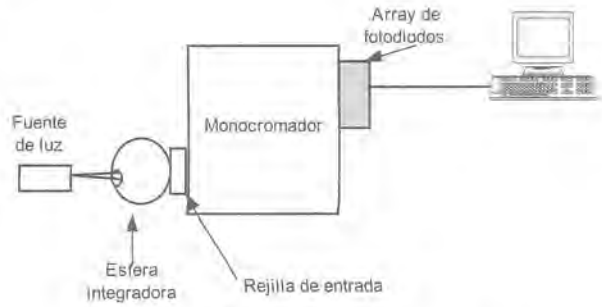


Figura 19.51. Espectro-radiómetro.

Otro instrumento muy utilizado es el **espectro-fotómetro** utilizado para determinar las propiedades de transmitancia y reflectancia de los materiales. Aunque se denomina espectrofotómetro porque su principal aplicación es la medida en el espectro visible, a menudo se diseñan para medir en UV o en IR. La diferencia con el anterior radica en que un espectro-radiómetro mide fuentes de luz que se sitúan externas al sistema, *mientras que un espectro-fotómetro incorpora fuentes internas que inciden sobre la muestra a estudio. La* mayoría de los espectro-radiómetros y espectro-fotómetros utilizan como elementos sensores un conjunto de fotodiodos.

Radiómetros y fotómetros

Los **radiómetros** se utilizan para medir la potencia radiada sobre un amplio margen de longitudes de onda. Consisten básicamente en el sensor con su circuito de acondicionamiento y un sistema óptico para recoger la

[20] Estas esferas se utilizan como fuentes de radiancia uniforme y se usan a menudo para la medida del flujo luminoso de una determinada fuente que se sitúe en su interior. La esfera captura la luz emitida por la fuente en todas las direcciones.

radiación de la fuente que pretende medirse. Los radiómetros se suelen utilizar con el detector y el filtro adecuado al margen de frecuencias que se pretenda medir. El filtro debe compensar la selectividad espectral del sensor y las características de transmisión del sistema óptico, buscando una respuesta más o menos plana en el detector para el margen de longitudes de onda a medir. Un radiómetro con un filtro, de manera que la respuesta espectral del sensor se aproxime a la respuesta espectral del ojo humano, se denomina **fotómetro**.

19.4.2 SENSORES ÓPTICOS

En general los sistemas de detección ópticos utilizan dos tipos de sensores:

- **Sensores térmicos:** termopilas y piroeléctricos.
- **Sensores fotónicos:** fotomultiplicadores, fotodiodos y fotoconductores.

En la medida de radiación óptica (entre 100 nm y 10000 nm) la primera consideración a tener en cuenta es la sensibilidad del detector en la banda de interés. También debe considerarse su linealidad, su estabilidad y su durabilidad. Por supuesto, cada sensor presenta su peculiaridades; por ejemplo, algunos detectores necesitan refrigeración, en otros es necesario modular la radiación o requieren una alta tensión para funcionar, otros pueden dañarse si se someten a una excesiva radiación o se toca su ventana sensible, etc. Cada una de estas peculiaridades debe valorarse a la hora de seleccionar la mejor opción. En la figura 19.52 se representa los márgenes de detección de distintos tipos de sensores de radiación (sin filtrado) lo que constituye uno de los elementos de decisión más críticos. En general, podemos controlar el pico máximo de detección de un sensor y su ancho de banda por medio de filtros.

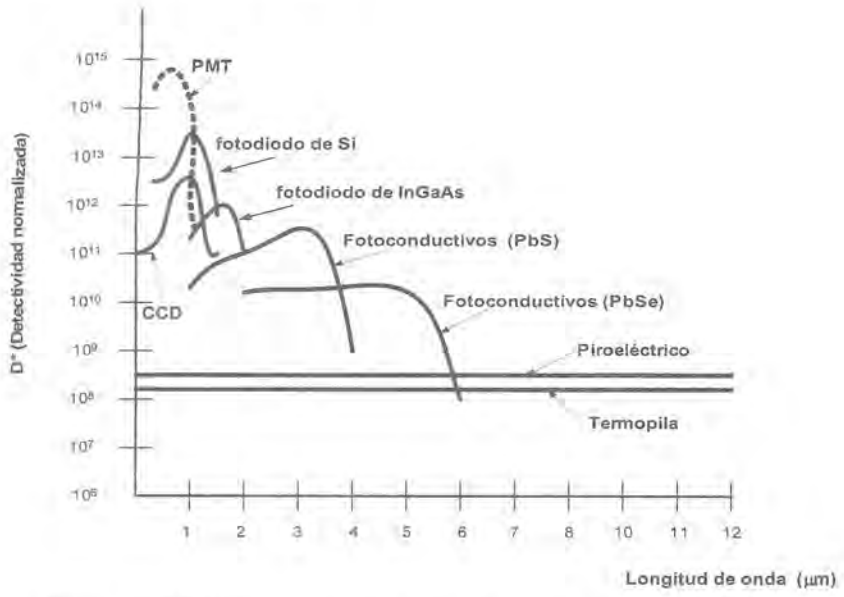


Figura 19.52. Márgenes de detección de sensores ópticos [19.24].

Figuras de mérito de los detectores ópticos

La mayoría de los detectores ópticos se describen por una serie de figuras de mérito que encontramos en sus hojas de datos. Estas figuras de mérito son función de su longitud de onda y de la temperatura, pudiendo

verse afectadas también por el tamaño del detector, la frecuencia de modulación y la tensión de polarización. Las figuras de mérito más importantes son:

- **Responsividad (S):** se define por $S = Y/X$; donde Y es la salida del detector y X es la entrada de radiación. Generalmente Y tiene un valor Y_0 para $X = 0$ (salida en oscuridad). En este caso: $S = (Y-Y_0)/X$. Normalmente, S depende de la longitud de onda. La variación con la longitud de onda se describe por la **responsividad espectral**: se define por $S(\lambda) = dY(\lambda)/dX(\lambda)$, donde $dY(\lambda)$ es la fracción de salida para la entrada $dX(\lambda)$ a la longitud de onda λ.
- **Eficiencia cuántica (η):** es el cociente entre el número de fotoelectrones producidos por el detector y el número de fotones incidentes.
- **Constante de tiempo (τ):** es el tiempo que emplea la salida del detector en alcanzar el 63% de su valor final ante un cambio en escalón de la fuente de excitación (constante de tiempo de subida). Cuando se elimina la fuente de excitación el tiempo hasta alcanzar el 36% de su valor inicial describe la constante de tiempo de bajada. Generalmente estos tiempos son diferentes.
- **Margen de linealidad:** define la diferencia entre el máximo y el mínimo flujo radiante de entrada para el cual la salida del detector es lineal.
- **Potencia equivalente de ruido (NEP):** describe el flujo radiante necesario para obtener a la salida un valor igual al valor eficaz de ruido generado en el propio sensor.
- **Detectividad (D):** es el recíproco de NEP. Cuanto más sensible es el detector mayor es el valor de este parámetro.
- **Detectividad normalizada (D*):** permite una comparación entre diferentes tipos de detectores con independencia del área del detector y de su ancho de banda: $D^* = D(A^{1/2}\Delta f^{1/2}) = (A\Delta f)^{1/2}/NEP$, donde A es el área del detector y Δf su ancho de banda; sus unidades son $cmHz^{1/2}W^{-1}$.
- ***Ruido:*** *esta característica limita la detectividad de los sensores ópticos y pueden distinguirse varios* tipos (véase capítulo 2). Además de estas fuentes de ruido, los detectores pueden estar sujetos a ***ruido microfónico*** (*microphonic noise*) debido a vibraciones o impactos y el ***ruido de temperatura*** (debido a fluctuaciones en la temperatura, suele ser un problema en los sensores pequeños con baja inercia térmica). También es importante considerar el ruido debido a las etapas de amplificación y acondicionamiento de señal (**ruido de postdetección** y **ruido de digitalización**) a menudo los circuitos electrónicos después de la detección determinan los niveles de señal más pequeños medibles, particularmente en los detectores que no incorporan internamente ningún tipo de amplificación.

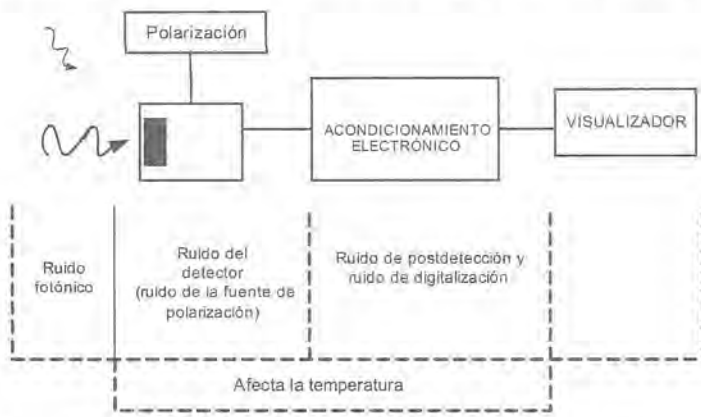


Figura 19.53. Ruidos en el sistema de detección [19.24].

Detectores térmicos

Los detectores térmicos convierten la radiación incidente en un incremento de temperatura en el sensor. El incremento de temperatura produce un efecto en el sensor; por ejemplo, se genera una tensión (termopares) o se induce una carga (piroeléctricos). Generalmente la superficie del sensor se trata depositando una sustancia que absorbe de forma óptima y uniforme la radiación deseada lo que contribuye a incrementar la responsividad del sensor y su margen espectral de medida; esta es una gran ventaja de este tipo de sensores.

Termopilas

Las termopilas están constituidas por un conjunto de termopares conectados en serie (de 20 a 120) con objeto de incrementar la tensión de salida y con todas las uniones muy juntas con objeto de recoger toda la radiación (véase capítulo 13). Una característica importante es que pueden fabricarse termopilas a muy bajo coste, como ocurre, por ejemplo, con las termopilas de película delgada cuyo tamaño también puede hacerse muy pequeño. La constante de tiempo en las termopilas miniatura puede ser muy reducida (hasta 50 ms). Las termopilas no presentan ruido de *flicker*, 1/f, debido a que no es necesario alimentarlas (son sensores generadores). Pueden realizarse medidas con gran sensibilidad desde continua hasta algunos hertzios fundamentalmente en el infrarrojo. Las medidas suelen hacerse troceando la radiación que incide sobre el detector y observando el cambio en la tensión de salida.

Piroeléctricos

Están basados en un material cristalino dispuesto entre dos electrodos capaz de inducir una carga sobre los electrodos por efecto térmico (véase capítulo 15). Estos sensores responden sólo a la radiación pulsada o modulada que es la que provoca el cambio en la temperatura del material. Responden de forma más rápida que los termopares o termopilas, y además no les afecta el ruido permanente de radiación de fondo. Pueden fabricarse sensores piroeléctricos de tamaños muy reducidos y, por tanto, con respuestas muy rápidas.

Detectores fotónicos o cuánticos

Estos sensores responden al flujo fotónico de forma cuantificada. La respuesta del sensor puede ser la generación de un fotoelectrón en un sensor que contenga una sustancia fotoemisiva (como el fotomultiplicador), la generación de una pareja electrón-hueco en una unión semiconductora (fotodiodos), o el paso de un electrón de la banda de valencia a la de conducción (fotoconductor). El resultado en cualquiera de los casos es una corriente o una tensión relacionada con la fuente de radiación.

Las características básicas de los detectores fotoemisivos y los basados en semiconductores de unión (fotodiodos) pueden resumirse en [19.24]:

1. La interacción electrón-fotón es muy rápida, así que estos sensores responden de forma muy rápida a flujos de radiación cambiantes.

2. Su detectividad es, en general, más alta que la de los detectores térmicos, aunque sobre un margen más limitado de longitudes de onda.

3. Todos los mecanismos de detección dependen de la longitud de onda, es decir, hay un pico de responsividad a una longitud de onda corta y otro a longitud de onda larga (figura 19.54). A longitud de onda larga (baja energía en los fotones) ocurre el decaimiento de la responsividad debido a que se requiere una mínima energía para provocar la fotoemisión o generación de parejas electrón-hueco. La baja responsividad a longitud de onda corta se debe a dos causas: hay menor número de fotones con

longitud de onda corta por vatio[21], y por otro lado a longitudes de onda cortas los fotones no pueden ser absorbidos adecuadamente por la región sensitiva.

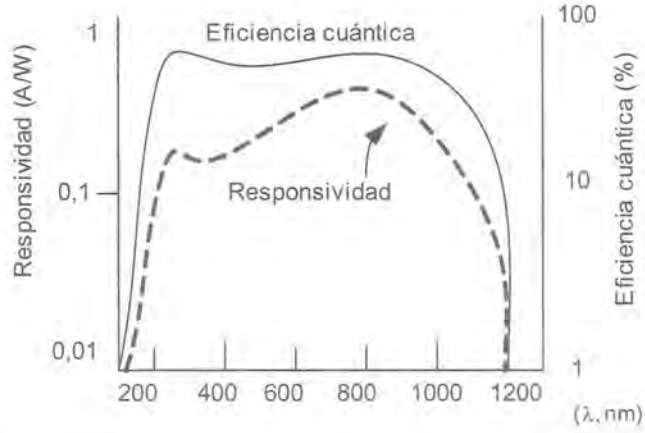


Figura 19.54. Responsividad y eficiencia cuántica de los detectores fotónicos [19.24].

[21] La energía del fotón es inversamente proporcional a su longitud de onda, así, por ejemplo, para la misma potencia de radiación, el flujo de fotones es más bajo en el ultravioleta que en el infrarrojo.

20

Introducción a la transmisión de señal

20.1 La transmisión
20.2 Telemedida
20.3 Bucles de tensión y bucles de corriente

20.1 La transmisión

En algunas ocasiones, el lugar donde se genera la información de las variables de un sistema de instrumentación está separado una cierta distancia del lugar en que esa información va a ser utilizada; por ejemplo, es frecuente que en las plantas industriales los procesos de fabricación cuyas variables deban ser medidas se encuentren en un determinado edificio y los centros de control, en otro. Esto puede hacerse extensivo a otras aplicaciones que van desde ejemplos más o menos simples de domótica hasta sistemas espaciales.

En todos estos casos, es necesario hacer que la información recorra distancias que resultarán más o menos grandes en un proceso genérico que puede denominarse "transmisión de información".

En toda transmisión se debe distinguir entre lo que son *datos* y lo que son *señales*. Se entiende por dato cualquier entidad que contiene únicamente información y que, por ello, constituyen el núcleo del proceso de transmisión. La forma en que los datos se configuran para ser transmitidos desde un origen hasta un destino es lo que se denomina señal y puede adoptar diversas naturalezas en función de la vía que se emplee para enviar esa señal. Tanto los datos como las señales pueden ser analógicos o digitales y se pueden dar, en principio, las cuatro combinaciones posibles:

a) Señales analógicas que contienen datos digitales.
b) Señales analógicas que contienen datos analógicos.
c) Señales digitales que contienen datos digitales.
d) Señales digitales que contienen datos analógicos.

En la figura 20.1 se muestra un esquema general de un proceso de transmisión.

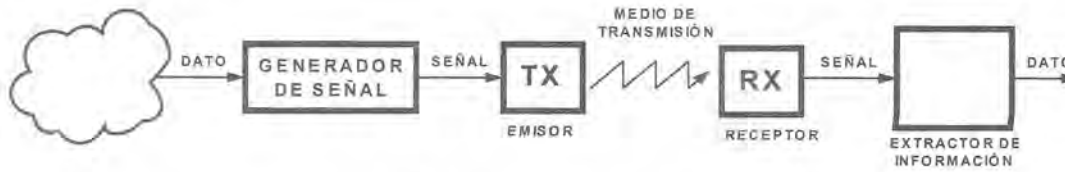


Figura 20.1. Diagrama de bloques de un sistema de transmisión de información genérico.

Como se puede apreciar, además de datos y señales y los correspondientes sistemas que generan las señales o extraen los datos, hay que destacar como importante elemento del proceso de transmisión el camino por el que discurren las señales desde el emisor hasta el receptor: el *medio de transmisión*.

La propagación de las señales por el medio de transmisión lleva implícito un cierto riesgo puesto que es posible que sufran perturbaciones por el camino procedentes de cualquier causa. En principio, las perturbaciones que va a sufrir la señal en el medio no van a ser diferentes de las que sufrirá cualquier señal eléctrica en un circuito o sistema electrónico y, por ello, se podrán estudiar igual que se comentaba en los primeros capítulos. También los efectos que se producirán van a ser similares a los que se tenían entonces ya que redundarán en poner en peligro la información contenida en las señales.

Como causas de perturbación tenemos las producidas por el propio medio y que afectan a la señal por el mero hecho de propagarse por él (causas propias) y las que están producidas por otros agentes presentes en el medio o que lo utilizan (causas externas); así tenemos:

a) Causas propias: básicamente son los efectos de atenuación y de distorsión producidos por considerar que la propagación no es ideal, sino que supone un camino con pérdidas y con limitación de ancho de banda como cualquier otro sistema físico.

b) Causas externas: fundamentalmente son los ruidos y las interferencias con otras señales presentes en el medio.

El efecto producido se traduce en una pérdida de señal y un incremento de ruido que se cuantifica mediante la relación S/N o, más genéricamente mediante el término de THD+N (véase capítulo 2).

En el peor caso, se podría llegar a deformar tanto la señal que se pierda la información contenida con lo que se ocasionaría una rotura en la transmisión.

En conclusión, las señales se propagan por el medio de transmisión y actúan como soporte-contenedor de la información que se desea transmitir preservando la información en ellas contenida.

20.1.1 MEDIOS DE TRANSMISIÓN

El medio de transmisión constituye el camino físico entre el sistema de emisión y el de recepción según se muestra en la figura 20.1: el emisor introduce una señal adecuada para su propagación "fácil" en el medio y el receptor recogerá esa señal del medio para llevarla hasta el sistema que extraiga la información que contiene.

El medio de transmisión, como camino para la señal, debe permitir su propagación cómoda lo que se traduce en: no atenuar, no distorsionar y no introducir ruido. Claro está que un medio capaz de comportarse así sería un medio "ideal" y, como es fácil de suponer, no existe. Sin embargo, el objetivo del medio de transmisión será parecerse lo más posible al ideal establecido introduciendo pocas perturbaciones en la señal original.

Cualquier medio físico puede ser utilizado para la transmisión de señales aunque, a grandes rasgos se puede hablar de:

a) Medios guiados que suponen el confinamiento de las señales en el medio por el cual se propagan. Suele tratarse de medios con soporte sólido tales como cables eléctricos o fibras ópticas.

b) Medios no guiados o libres en los que el soporte es ilimitado o cuasi-ilimitado y las señales evolucionan "libremente" por el medio. Medios de este tipo son el aire, el agua o el vacío.

En el primer caso, el confinamiento de la señal implica una cierta protección y las interferencias externas al medio afectarán más difícilmente a la señal que se propaga; además la potencia que supone la señal de propagación no se dispersará (o lo hará poco) de tal manera que presentarán poca atenuación.

En el segundo caso, la libertad intrínseca del medio no pone trabas a las perturbaciones externas y serán posibles las interferencias por lo que la señal no queda demasiado protegida; más aún, la libertad de propagación dificulta el encaminamiento de la energía en una determinada dirección con lo que las señales tienden a dispersarse y, consecuentemente, sólo una pequeña parte de la energía que emite el emisor llegará al receptor introduciendo una fuerte atenuación en el proceso de transmisión.

Aunque podría parecer que los medios guiados siempre son mejores que los libres, los primeros obligan a un nexo físico entre emisor y receptor lo que limita su utilización; esta limitación no existe en el otro caso dotándolo, en general, de una gran flexibilidad.

20.1.1.1 Medios de transmisión guiados

Aunque serían posibles otros soportes, en el ámbito de la transmisión de señales se puede hablar de cables de cobre y de cables de fibra óptica como medios de transmisión en los que la señal va confinada en el medio.

Por los cables de cobre se propagarán señales de naturaleza eléctrica, es decir, señales de tensión o de corriente producidas por un emisor y que alcanzarán al receptor si la energía que llevan en el momento de la llegada es suficiente como para superar el mínimo valor necesario. La velocidad de propagación por los cables es la de la luz, que puede parecer un valor muy elevado pero que, en realidad, no lo es tanto en la escala de tiempos que se maneja en electrónica; así, introducirá retrasos que pueden llegar a ser apreciables en función de la distancia del enlace.

Por los cables de fibra óptica viajan señales de luz introducidas por el transmisor (algún tipo de LED o diodo láser) y que, tras hacer el trayecto, serán recogidas por el receptor (algún tipo de fotodiodo). Los retardos impuestos por la propagación de la onda son similares a los del caso anterior y tendrán o no importancia en función de la distancia del enlace.

Aunque hacer un estudio completo del proceso de transmisión a través de este tipo de medios tiene un indudable interés, se escapa del ámbito de este libro, entrando dentro del campo de la ingeniería de telecomunicación por lo que nos restringiremos a dar una breve información sobre estos aspectos, dejando para la curiosidad del lector una mayor profundización.

Cables de cobre

Como se comentó, el *hilo de cobre* puede servir para llevar señales de corriente o tensión de un emisor a un receptor; para ello, se pueden tener diversas configuraciones geométricas que presentan más o menos ventajas; para comprender por qué unas opciones son mejores que otras hay que tener en cuenta los aspectos negativos que afectan al proceso de transmisión: atenuación, ancho de banda e interferencias.

La solución más sencilla es la de un hilo conductor con retorno por tierra (figura 20.2a) que se usó en los primeros sistemas telegráficos pero que presenta graves problemas de ruido ocasionados por las interferencias.

La utilización de dos *hilos paralelos* representa una mejora respecto de la solución anterior —aunque duplica el coste— por cuanto es más resistente a las interferencias externas (figura 20.2b) ya que el área encerrada por la espira que forma el sistema es menor y, por ello, serán menores las señales inducidas.

Un paso adicional que permite disminuir más aún el efecto de las interferencias externas consiste en el trenzado de los hilos (*par trenzado*) que permite que las señales interferentes sean aún menores y, además, se anulen entre sí en una buena medida (figura 20.2c). Esta última es la solución que más se suele adoptar para el envío de señales en el entorno de la instrumentación electrónica.

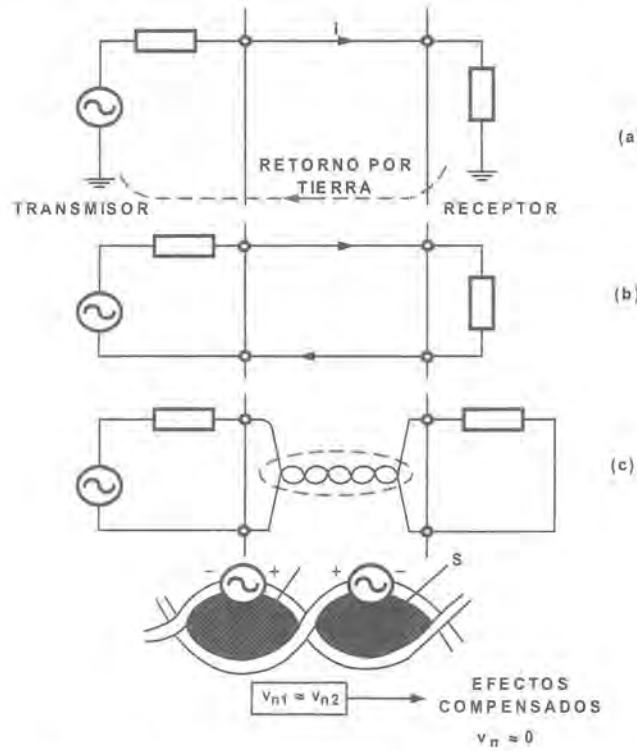


Figura 20.2. Uso de cables de cobre para transmisión de señales: (a) un hilo y retorno por tierra; (b) dos hilos paralelos; (c) dos hilos trenzados "compensan" la generación de ruido.

El cable apantallado puede mejorar su comportamiento ante las interferencias mediante la inclusión de una pantalla metálica exterior que rodea a los hilos por los que se propaga la señal. Se obtiene así el par

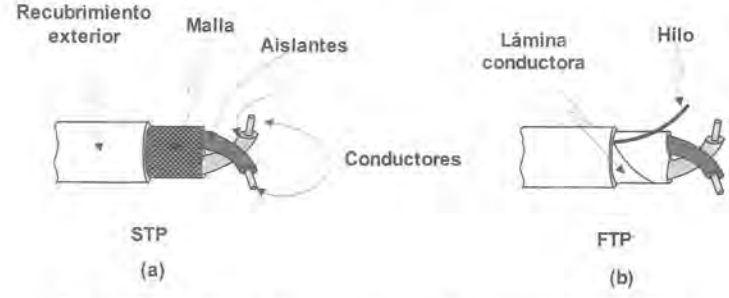


Figura 20.3. Tipos de pares trenzados: (a) STP; (b) FTP.

trenzado apantallado (STP, de la terminología anglosajona *Shielded Twisted Pair*) en contraposición con el UTP anterior (*Unshielded Twisted Pair*). El STP incluye una malla metálica rodeando todos los cables (figura 20.3a) y puede encontrarse en una variante mejorada, el FTP (*Foiled Twisted Pair*) en el que la pantalla es una lámina e incluye un hilo que garantiza su continuidad (figura 20.3b).

En la tabla 20.1 se indican algunos tipos de pares trenzados de 100 Ω con sus aplicaciones típicas.

Tipo	Número de pares	Aplicaciones
UTP	1 ó 2	Telefonía
	4	Cable estructurado
	10, 20, 25, 50 100, 300	Troncales de telefonía Datos baja velocidad Datos velocidad media Instrumentación
STP	1, 2, 4, 25	Instrumentación
FTP	4	Cable estructurado
	25	Telefonía Datos Instrumentación

Tabla 20.1. Tipos de pares trenzados y sus aplicaciones.

El cable de tipo FTP es más recomendable que los otros tipos pero es algo más difícil de manejar (conectorizar) y es más costoso. En general, cuando se trabaja con cables con varios pares trenzados, el proceso de conectorizado puede llegar a ser muy laborioso y, en cualquier caso, debe realizarse con cuidado, recordando que los hilos del mismo par se identifican porque uno es de un determinado color y su correspondiente es blanco con líneas de ese color.

En algunos casos, se utilizan cables de tipo coaxial con dos conductores concéntricos que son especialmente interesantes por la menor atenuación respecto de los anteriores, su mejor respuesta en frecuencia y una aceptable inmunidad al ruido interferente (figura 20.4); este tipo de cables se suele utilizar en aplicaciones de radiofrecuencia y en algunos casos de redes locales. La atenuación que introduce este tipo de cable es la determinada por la expresión [20.9]:

$$A_t = k\frac{\sqrt{f}}{\log(D/d)}\left(1/D + 1/d\right)) \qquad (20.1)$$

donde k es una constante que depende del aislante del conductor interior y que hace de dieléctrico entre ambos conductores, D es el diámetro interior del tubo y d, el del conductor central. Como se desprende de la expresión 20.1, la geometría del cable es muy importante en el parámetro atenuación pero, más aún, para las aplicaciones de este tipo de cables, resulta muy importante el valor de su impedancia característica que queda determinada por la relación entre los diámetros de los dos conductores que configuran el cable coaxial y que resulta un parámetro crítico a la hora de evitar rebotes en el cable cuando este se ve como una línea de transmisión [20.6].

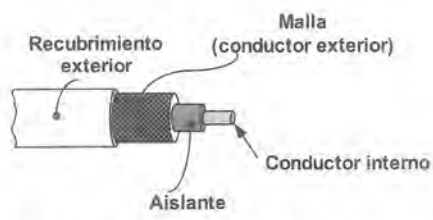


Figura 20.4. Cable de cobre coaxial.

Cables de fibra óptica

La fibra óptica, que pasó casi desapercibida durante los años que siguieron a su aparición y fue relegada a aplicaciones anecdóticas ha cobrado un protagonismo total en el campo de la comunicación y, hoy en día, se ha convertido en el medio de transmisión de información por excelencia. Su aplicación también se ha extendido a campos afines como el de la electrónica de potencia y el de la instrumentación electrónica gracias a sus excelentes prestaciones a la hora de salvaguardar la información ante perturbaciones externas. Incluso, en algunos casos, se ha convertido en un sensor en sí misma [20.2], campo en el que aún está en un estado inicial aunque se han conseguido ya algunos dispositivos comerciales.

El funcionamiento de la fibra óptica como medio de transmisión se basa en el confinamiento de la señal de luz en un medio gracias a la diferencia entre el índice de refracción de ese medio (*núcleo*) y el índice de refracción del medio que lo rodea (*corteza*) según se indicó en el apartado 18.1. Las fibras ópticas de "salto de índice" se suelen nominar por los diámetros de núcleo y corteza: por ejemplo, se habla de fibra de 125/62,5 como fibra de 62,5 micras de diámetro de núcleo y 125 micras de diámetro de corteza.

El mismo efecto se puede conseguir haciendo variar paulatinamente el índice de refracción desde n_1 hasta n_2 como se indica en la figura 20.5. El efecto que se apreciaría sería el de un haz que se va curvando poco a poco hasta retornar a la zona más interna debido a un proceso de reflexión-refracción diferencial.

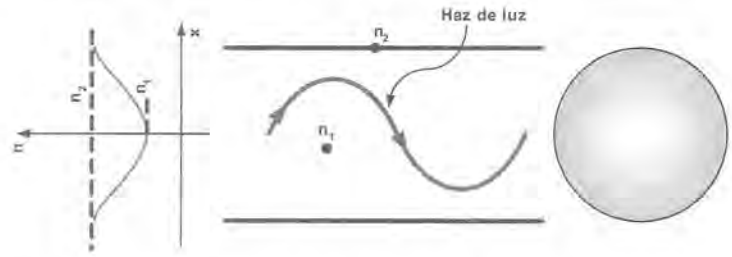


Figura 20.5. Propagación de la luz en una fibra de índice gradual.

En el caso de la fibra de la figura 20.5 estamos en presencia de una *fibra óptica de índice gradual*. Esta fibra sólo se nomina por su diámetro externo.

La luz en el interior de la fibra no se puede propagar por cualquier camino, sino que sólo es posible un determinado número de ellos; cada uno de estos "caminos de propagación" se denominan modos y, según sea su número se puede clasificar la fibra entre *fibra monomodo* (*singlemode*) cuando sólo es posible uno de ellos o *fibra multimodo* (*multimode*) cuando son posibles varios (véase apartado 18.1) [20.5].

Las fibras ópticas se construyen con materiales que van desde el plástico a diversos materiales cristalinos como el cuarzo o similares al cuarzo. La fibra óptica se presenta en cables de una o varias fibras y suele incluir elementos adicionales que le proporcionan un mejor comportamiento mecánico y/o mayor inmunidad frente a las hostilidades exteriores. En la figura 20.6 se muestra un tipo de cable de fibra óptica con cuatro fibras.

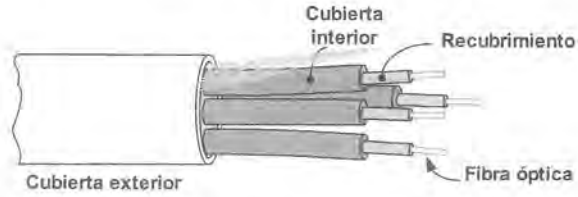


Figura 20.6. Cable de fibra óptica de cuatro fibras.

La fibra óptica debe ser conectorizada en sus extremos para un manejo razonablemente cómodo. Aunque este proceso no está exento de dificultad, existen algunas soluciones con diversos tipos de conectores tanto para fibra de plástico como para diversas variedades de fibra de cuarzo. Por ejemplo, tenemos conectores metálicos de tipo SMA para fibras desde 100 micras hasta 1,5 mm mientras que conectores de tipo ST o SC sólo se usan para fibras de 100 ó 125 micras de diámetro.

Queda fuera del alcance de este libro una explicación pormenorizada sobre las diversas posibilidades y formas de trabajo con fibra óptica y, para ello, deberá recurrirse a libros más especializados.

Sin embargo, hablando en los términos que caracterizan un sistema de transmisión, es decir, las perturbaciones que van a surgir a lo largo del camino de propagación, tenemos que la fibra óptica constituye un medio que distorsiona muy poco la señal y que se ve muy poco afectado por interferencias exteriores. La poca distorsión sobre la señal se debe a su elevado ancho de banda y a su baja distorsión, y la poca influencia de los agentes interferentes exteriores se debe a la propia naturaleza de la luz que es poco sensible a la influencia de campos electromagnéticos.

Las prestaciones de atenuación y ancho de banda son dependientes en cada fibra y, en este sentido, la fibra plástica presenta las peores características, mientras que las fibras monomodo son las que proporcionan menores atenuaciones y mayor ancho de banda permitiendo enlaces de bastantes kilómetros sin mayor dificultad.

En muchos casos, las principales restricciones vienen impuestas por los sistemas que introducen la luz en el interior de la fibra (emisores) a partir de las señales eléctricas y los que en el extremo de llegada recuperan la luz obteniendo las mismas señales eléctricas de partida (receptores). En general, se utiliza algún tipo de diodo luminiscente, ya sea de tipo LED o de tipo láser (LD) para la emisión y fotodiodos de tipo PIN o de avalancha en la recepción (véase el capítulo 16); así, el conjunto aparece como el que se muestra en la figura 20.7.

Para salvaguardar las excelentes características del medio, es preciso que los dispositivos implicados en el proceso (emisor de luz, fotodetector y fibra óptica) estén perfectamente adaptados en espectro luminoso, es decir, que presenten un conjunto de longitudes de onda comunes en las que su comportamiento sea bueno. En general, podrían existir más zonas de trabajo y, de hecho, en algunas aplicaciones se usan las fibras para llevar luz de otros colores, pero lo más normal es que se empleen siempre con luz de tipo infrarrojo. Más en concreto, se suele trabajar en tres determinadas zonas del espectro, denominadas *ventanas de transmisión* y que corresponden a zonas alrededor de longitudes de onda de 850, 1350 y 1550 nm.

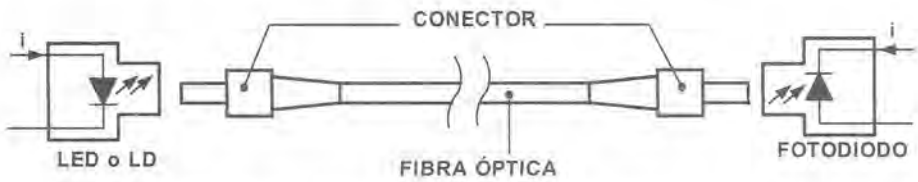


Figura 20.7. Cable de fibra óptica entre un emisor y un receptor.

Una vez seleccionada la ventana de transmisión y el resto de dispositivos —el emisor, el receptor y la fibra— más adecuados para nuestros objetivos, quedará definido el ancho de banda y el alcance del enlace: el ancho de banda quedará definido por el menor de los anchos de banda de los sistemas implicados en el proceso y el alcance, por la atenuación producida en el camino.

Para determinar el alcance de un enlace de fibra óptica se partirá de la potencia de luz generada por el emisor y se calculará el nivel de potencia que llega al receptor teniendo en cuenta las atenuaciones de las conexiones y de la propia fibra. Si el nivel de potencia luminosa es mayor que el valor de la potencia equiva-

lente de ruido (NEP)[1] del receptor, será posible la discriminación de la señal y el enlace trabajará correctamente; sin embargo, si el nivel de potencia es menor que el valor de la NEP, tendremos una relación S/N menor de cero en la recepción y la discriminación de la señal no será posible o implicará un tratamiento adicional de la señal recibida. Como quiera que la atenuación más importante suele ser la introducida por la propia fibra y esta depende de la longitud, para un conjunto emisor-fibra-receptor determinado, habrá un alcance máximo. En la figura 20.8 se muestra un diagrama de potencia de luz en un enlace de fibra óptica.

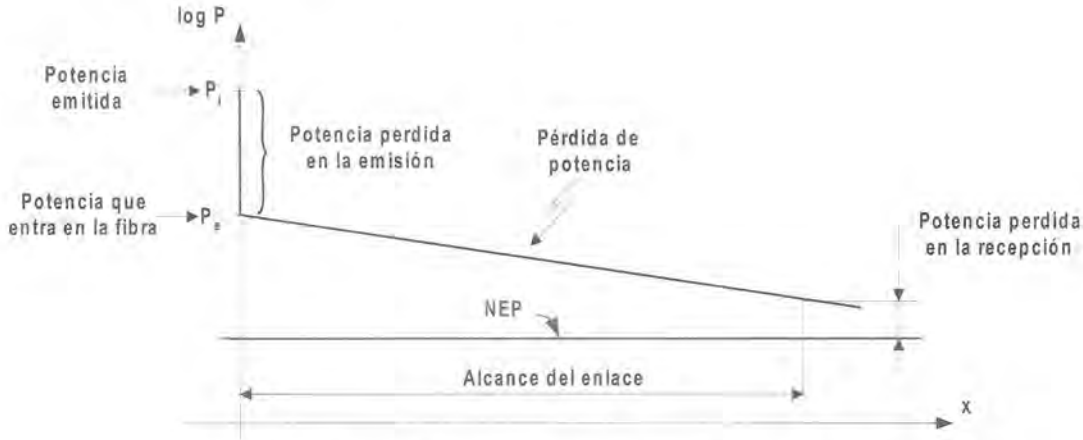


Figura 20.8. Pérdida de potencia en un enlace de fibra óptica.

Hay que indicar que de toda la potencia que genera el emisor, sólo una parte entrará en la fibra y, de toda la potencia que salga de la fibra, sólo un porcentaje llegará al receptor.

EJERCICIO 20.1

Supongamos que se desea realizar un enlace de fibra óptica en la primera ventana de transmisión utilizando un emisor de Honeywell del tipo HFM2010-224 capaz de proporcionar una potencia de 120 μW de luz de 820 nm a una fibra de 100 micras de diámetro, un receptor del mismo fabricante tipo HFM1011-221, una fibra óptica de National Instruments de tipo T3 con una atenuación de 4,3 dB/km a 820 nm y conectores adecuados (SMA 905) que introducen una atenuación máxima de 1,5 dB. ¿Cuál es el alcance máximo del enlace?

SOLUCIÓN:

En la figura 20.9 se muestra un diagrama de potencias indicando dónde se va perdiendo la señal.

La pérdida debida al primer conector es, de 1,5 dB, lo que significa que P_{C1} es:

$$1{,}5 \text{ dB} = 10\log\frac{120}{120 - P_{C1}} \Rightarrow P_{C1} = 35\,\mu\text{W}$$

con lo que llegará a la fibra una potencia de 85 μW.

[1] Muchos fabricantes proporcionan un valor genérico que indica la mínima potencia que el receptor es capaz de discriminar.

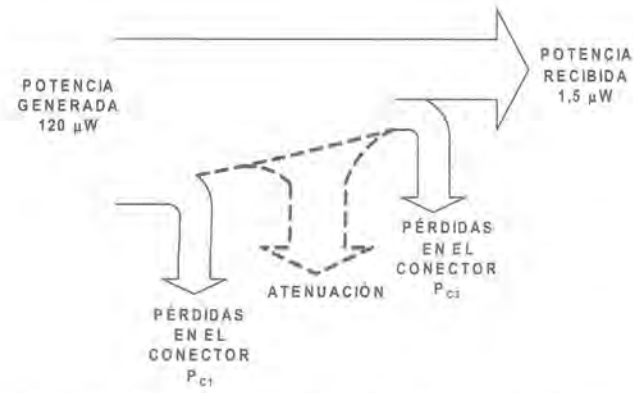


Figura 20.9. Pérdida de potencia en un enlace de fibra (los grosores de las flechas no están a escala).

En el receptor, la potencia mínima necesaria para que se pueda interpretar la señal es de 1,5 μW. Teniendo en cuenta las pérdidas del conector:

$$1{,}5 \ \text{dB} = 10\log\frac{1{,}5+P_{C2}}{1{,}5} \Rightarrow P_{C2} = 0{,}6 \ \ \mu\text{W}$$

se precisa que en el extremo de llegada de la fibra haya, al menos, 2,1 μW lo que implica que la señal de partida puede atenuarse en:

$$A_{\text{t, máxima}} = 10\log\frac{85}{2{,}1} = 16\,\text{dB}$$

Como quiera que la atenuación a la longitud de onda de trabajo es 4,3 dB/km, se podría conseguir un enlace a una distancia máxima de unos 3,7 km.

Lógicamente, no es conveniente apurar el tendido hasta los extremos del ejemplo porque cualquier circunstancia tal como un conector no del todo correcto, un doblez en la fibra o cualquier otra circunstancia adversa derivada de la Ley de Murphy, darían al traste con el funcionamiento. En estos casos, se suele dejar un cierto *margen de seguridad* que puede definirse en dB como el cociente entre la señal que llega al extremo y el valor de la NEP del receptor.

La transmisión de información por fibra óptica se suele hacer mediante señales digitales con lo que, para transmitir información analógica, habrá que efectuar un proceso de digitalización y codificación de esa información en la emisión e inverso en la recepción, tal y como se muestra en la figura 20.10.

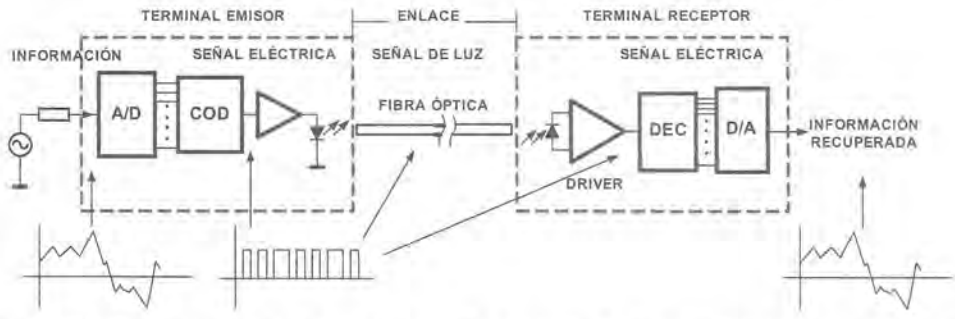


Figura 20.10. La transmisión de informaciones analógicas a través de fibra óptica se lleva a cabo a través de un proceso de digitalización.

En realidad, en los casos más frecuentes se suelen mezclar más señales en la emisión para aprovechar mejor el medio de transmisión con lo que los terminales emisor y receptor no son tan simples como los de la figura 20.10, apareciendo sistemas realmente complejos controlados por algún tipo de microprocesador. Además, se suelen incluir sistemas de modificación de la señal[2], inclusión de datos para sincronización, etc.

En estos casos, el terminal emisor suele estar constituido por el *driver* y el emisor de luz y el terminal receptor por el fotodetector y el circuito de conversión I/V correspondiente.

20.1.1.2 Medios no guiados

La principal ventaja de los medios guiados es la relativamente alta inmunidad de las señales que se propagan por ellos a las interferencias exteriores; sin embargo, la exigencia de un nexo físico entre el equipo que transmite y el que recibe limita su campo de aplicación a sistemas sin movilidad. Para el resto de los casos, en los que al menos uno de los sistemas tenga posibilidad de moverse y que ello implique la imposibilidad de nexo físico, se deberá utilizar un medio de transmisión no guiado en el que las señales se propagarán con una cierta libertad.

Se puede hablar de dos tipos de medios de transmisión no guiados según el tipo de ondas que se propaguen:

a) Ondas de presión en medios elásticos.
b) Radiación electromagnética en el aire o en el vacío.

Al primer caso, corresponden comunicaciones mediante ondas sonoras que no dejan de ser más que vibraciones mecánicas que se desplazan en medios materiales como el aire o el agua. Sistemas de este tipo ya se han comentado —para otras aplicaciones— en el caso de los ultrasonidos aunque quizá el caso más frecuente es el de la comunicación humana mediante voz y que está limitada a un alcance relativamente pequeño; en otros medios tales como el agua o como los sólidos, la propagación es mucho más fácil y se pueden tener alcances realmente muy grandes.

Sin embargo, cuando se habla de transmisión de señal en medios no guiados se suele hacer referencia al segundo tipo de ondas, las electromagnéticas que tienen ancho de banda mucho más grande y alcances también mucho mayores, además de no precisar un medio físico de soporte para su propagación. Cuando se habla de radiación electromagnética, estamos tratando de señales con un amplio espectro que van desde las ondas hertzianas hasta los rayos gamma pasando por microondas, radiación luminosa y rayos X. En la figura 20.11 se muestra el espectro de las radiaciones electromagnéticas con ondas que van desde muchos km hasta unos pocos fm.

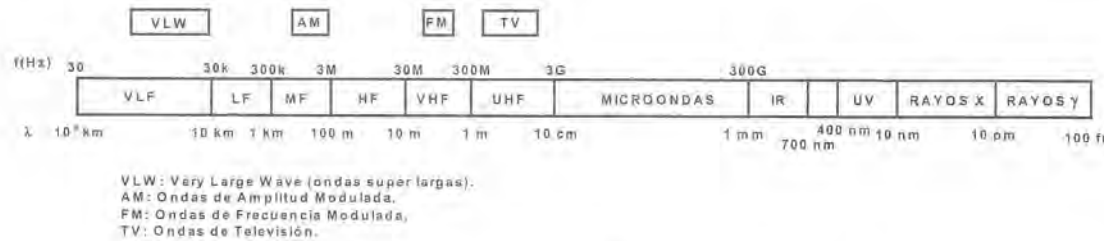


Figura 20.11. Espectro de ondas de radiaciones electromagnéticas.

[2] Además de la propia codificación de la información, la señal puede ser modificada para facilitar su lectura en el receptor e incrementar las garantías del proceso de transmisión.

La longitud de onda y la frecuencia están relacionadas por la expresión:

$$f = \frac{c}{\lambda} \tag{20.2}$$

donde f es la frecuencia, λ la longitud de onda y c la velocidad de la luz. Habitualmente, las ondas de radio y las microondas se suelen expresar en frecuencia, mientras que las demás radiaciones se indican con la longitud de onda. Desde el punto de vista de la transmisión de información, se usan básicamente ondas de radio en frecuencias elevadas y ondas de luz, aunque existen casos especiales como el uso de ondas superlargas de algunas decenas de km que tienen gran interés por su menor atenuación en medios como el agua.

Como unas y otras se propagan por medios no controlados, existen muchas posibilidades de que las señales sufran perturbaciones y, consecuentemente, son mucho menos seguros que los medios no guiados por lo que la información debe ser bien protegida en las señales que se propagan para que no se ponga en peligro la entidad de la misma.

Ondas de radio

La transmisión de señales mediante ondas de radio se realiza mediante un equipo transmisor que deposita la potencia sobre una antena capaz de ponerla en el medio y cuyas dimensiones son comparables a la longitud de onda de la señal que se propaga (esta es una de las razones por las que se tiende a utilizar frecuencia elevada). El receptor, mediante otra antena, recogerá la señal del medio recuperando una señal eléctrica copia de la original (figura 20.12).

Como el caso anterior, el problema del diseño y selección de las antenas, de los equipos de transmisión y recepción queda fuera del alcance de este libro ya que el número de factores implicados en el tema resulta en un problema complejo.

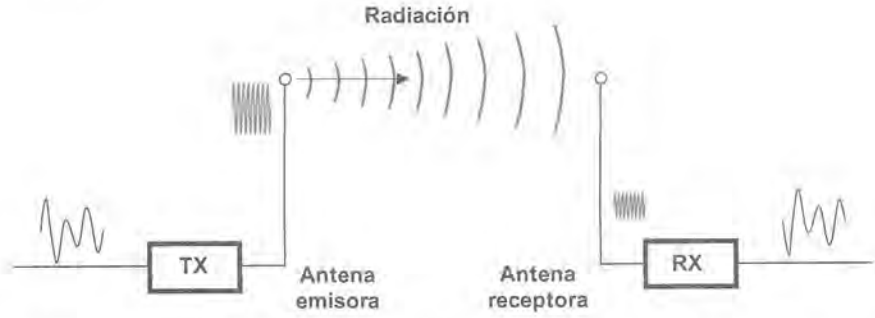


Figura 20.12. Esquema básico de un sistema de transmisión de información por radio.

En principio, la propagación de la señal desde la antena emisora hasta la receptora se realiza en línea recta aunque una buena parte de la energía radiante de salida se perderá y no alcanzará su destino; sin embargo, es posible emplear caminos no rectos entre emisor y receptor mediante la reflexión de las señales de radiofrecuencia[3].

[3] Un ejemplo típico es la reflexión de las señales de radio en la ionosfera que posibilitó la primera comunicación interoceánica. En este caso se realizó una transmisión entre Londres y Nueva York; sin saberlo, Hertz tuvo suerte y empleó una frecuencia de transmisión para la cual la ionosfera se comporta como un espejo permitiendo el rebote de la señal en ella y que, finalmente, alcanzase su objetivo. En algunas transmisiones se puede emplear el reflejo en la Luna aunque esto implique caminos muy largos con los consiguientes retrasos.

A la hora de utilizar un enlace de radio para la transmisión de cualquier tipo de señales hay que tener en cuenta factores adicionales tales como el ancho de banda que es mayor en la medida que lo sea la frecuencia de la señal que se transmite y las limitaciones impuestas sobre el espacio radioeléctrico que no permiten un uso indiscriminado del mismo: en efecto, el espacio radioeléctrico es común para todos los posibles emisores por lo que debe regularse de alguna forma para evitar que unas señales se solapen con otras convirtiéndolo en una especie de "jaula de grillos" en el que sólo se oye al que más grita. Para evitar que esto ocurra y que sea imposible utilizarlo por nadie, los organismos oficiales controlan el uso de las diversas *bandas de frecuencia* ya sea mediante la concesión de licencias, mediante la limitación de la máxima potencia permitida para las señales o evitando la utilización de determinadas zonas del espectro.

Cuando se pretenda emplear un enlace de radio con el fin que sea es conveniente consultar con la administración correspondiente para ver la disponibilidad de una banda y conocer los requisitos necesarios para poder utilizarla [20.7].

Ondas de luz

Uno de los primeros sistemas de telecomunicación empleados por el ser humano —al margen de la propia voz— fue la utilización de señales de luz[4]; las limitaciones que introduce este medio son bastante grandes ya que exige una visión directa emisor-receptor, la atmósfera no es demasiado transparente a estas longitudes de onda y, además las condiciones de humedad elevan fuertemente la atenuación.

Aun con estas limitaciones, existen algunos sistemas que permiten la comunicación en casos muy concretos:

a) Sistemas de tipo difuso en los que se emite luz de forma casi omnidireccional y que permiten cubrir una zona pequeña (algunos metros). Se usan sistemas de este tipo para enlaces en salas limitadas, con obstáculos variables y en los que se pretenda comunicar uno o varios sistemas móviles entre sí o con un sistema fijo. Un ejemplo de este tipo de comunicación es la comunicación con robots autoguiados o con vehículos guiados automáticamente en plantas de fabricación; la implementación de este tipo de sistemas implica la presencia de un gran número de emisores y receptores en los equipos que deben comunicarse (figura 20.13a).

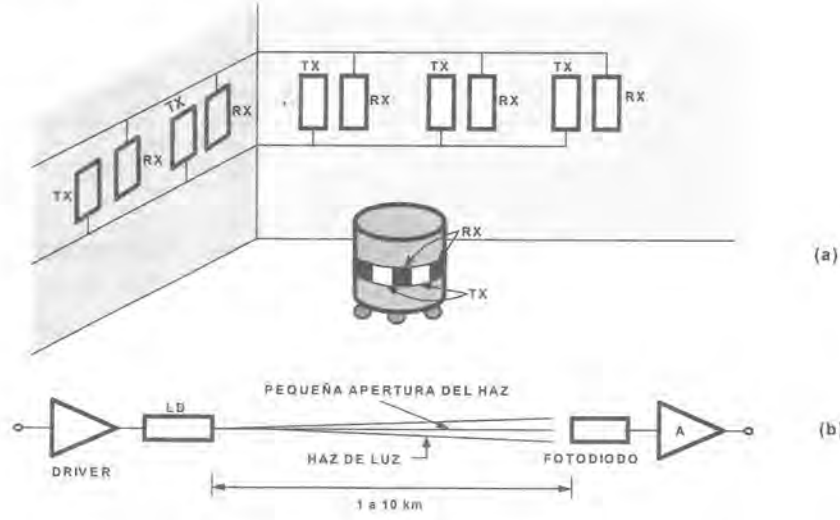


Figura 20.13. Transmisión de señal mediante luz en casos omnidireccionales (a) y unidireccionales (b).

[4] En la Era Moderna, el primer sistema fue desarrollado en Francia por Chappe que consistía en una serie de torres que iban transmitiendo la información de una a otra y que era capaz de cubrir distancias relativamente grandes.

b) Sistemas de tipo punto-punto en los que se emite luz unidireccional, normalmente mediante un diodo láser, que señala a un receptor determinado (figura 20.13b). El alcance de este tipo de enlaces puede ser de algunos (pocos) kilómetros aunque el volumen de datos que se transmite suele ser bastante alto. Técnicas de este tipo se emplean en plantas industriales con naves muy grandes en las que no sea fácil tender cable o el coste no lo permita. También se pueden emplear para comunicar con equipos móviles que se desplacen sobre raíles en línea recta como es el caso de los transelevadores de las plantas de almacenaje automatizado.

En todos los casos anteriores se utiliza la luz infrarroja que, además de pasar oculta a los ojos, presenta una menor atenuación en la atmósfera que la visible o la ultravioleta.

20.1.2 CODIFICACIÓN DE LA INFORMACIÓN

Con la información que se quiere enviar desde el emisor al receptor pueden tomarse dos acciones distintas:

a) Situarla tal cual en el medio de transmisión, efectuando el cambio a la magnitud adecuada para su propagación por él; en este caso, la señal que se propaga por el medio es una imagen directa de la señal de información (figura 20.14a).

b) Utilizarla para modificar una señal intermedia que luego será convertida a la magnitud más apropiada para su propagación por el medio; ahora la señal que se propaga no es una copia de la señal de información sino de otra señal intermedia que contiene —a su vez— la información (figura 20.14b).

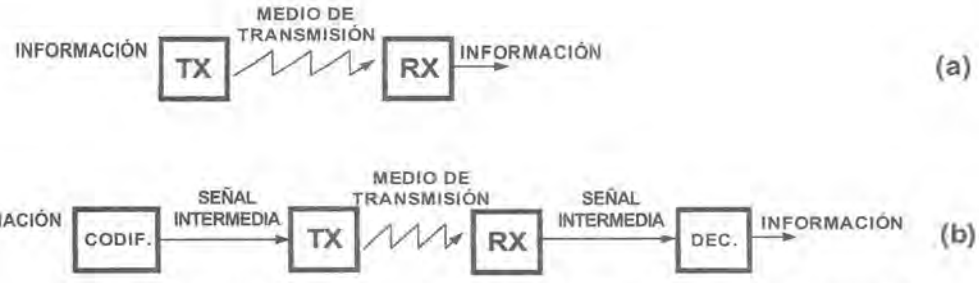


Figura 20.14. Transmisión de la información de forma no codificada (a) y codificada (b).

En el primer caso, cualquier efecto sobre la señal que se propaga se traducirá directamente en un cambio que puede afectar directamente a la información, mientras que en el segundo caso sería posible que los cambios afectasen a la señal que se propaga pero no a la información que esta contiene. En el primer caso, se habla de transmisión de la información en *banda base* mientras que, en el segundo, se habla de *modulación*.

Teniendo en cuenta las limitaciones impuestas por las perturbaciones producidas en el proceso de transmisión, el envío de la información en banda base sólo será factible cuando se tengan acotadas esas perturbaciones lo que sólo es posible en medios guiados. Por el contrario, en medios no guiados, lo imprevisible e incontrolable de las perturbaciones —sobre todo de las interferencias— hace que la información tenga que ser salvaguardada en otra señal. No obstante, aun en algunos casos de medios guiados tales como los cables de cobre, las perturbaciones pueden ser lo suficientemente intensas como para que también sea recomendable emplear alguna técnica de modulación.

El interés del empleo de la modulación, las ventajas que aporta y los inconvenientes serán tratados con más detalle en el siguiente capítulo y ahora nos restringiremos a asociar los medios de transmisión con las posibilidades de modulación según se muestra en la tabla 20.2.

Tipo de medio	Medio de transmisión	Alcance del enlace	Interferencias	Método de transmisión
Guiado	Cable de cobre	Corto	Bajas-medias	Banda base
			Altas	Modulación
		Medio	Bajas	Banda base
			Medias-altas	Modulación
	Fibra óptica	Cualquiera	Cualesquiera	Banda base
No guiado	Ondas de radio	Cualquiera	Cualesquiera	Modulación
	Luz	Cualquiera	Cualesquiera	Modulación

Tabla 20.2. Resumen general de los diversos medios de transmisión.

20.2 Telemedida

En muchos de los entornos de instrumentación el lugar en el que se debe efectuar la medición de un determinado parámetro o variable resulta definido por el propio proceso o sistema que desea ser monitorizado y suele ser inamovible. Sin embargo, el procesamiento de esa información hasta convertirla en un valor que será presentado, representado o, sencillamente, utilizado por un sistema de control puede tener lugar en otro lugar diferente del de medición y la distancia que los separe puede ser tan grande como el lector sea capaz de imaginar.

En las plantas industriales podemos encontrar muchos ejemplos de este tipo en que los procesos de fabricación/producción se llevan a cabo en naves acondicionadas para soportar los esfuerzos que supongan, mientras que el control de los mismos se ejecuta desde computadores en salas más o menos alejadas del proceso. En sistemas de supervisión de contaminación, en subestaciones eléctricas y en tantos otros lugares, se efectúan las medidas en un determinado punto y se llevan los valores hasta lugares que pueden estar a muchos kilómetros de distancia. En el extremo, cuando una sonda espacial o cualquier otra nave no tripulada toma una muestra, efectúa la lectura del contenido de cualquier gas en la atmósfera de un planeta y envía el resultado hasta el centro de control de la misión en la Tierra, se está llevando a cabo un proceso de medida a larga distancia.

En todos estos casos, el proceso de medida de parámetros a larga distancia que incluya la transmisión de la información captada por un sensor hasta un sistema receptor se conoce como *telemedida*.

El proceso de telemedida supone la presencia de tres grandes bloques: el sistema presente en el lado de emisión, el sistema de transmisión y el sistema presente en el lado de recepción (figura 20.15). Se puede llevar a cabo de muchas formas aunque, prescindiendo del propio medio de transmisión que se emplee, tenemos:

a) Telemedida analógica: todas las fases del proceso desde la lectura de la variable hasta la llegada al sistema que recibe la información son analógicas.

b) Telemedida digital: la señal se transforma en digital en el sistema del lado de emisión y se envía una información digitalizada por el medio de transmisión. En realidad, podría pensarse en digitalizar después de recibida la señal pero, teniendo en cuenta que las señales digitales son mucho más inmunes a las perturbaciones que las analógicas, cuanto antes se dé el paso, antes se reducirán sus efectos.

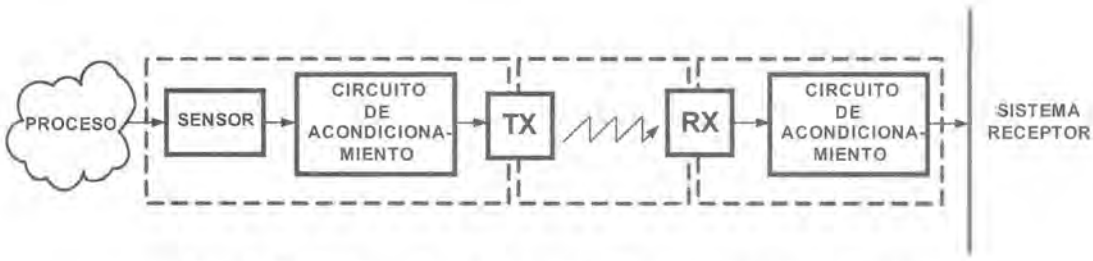


Figura 20.15. Sistema de telemedida como un concepto general.

La telemedida es un concepto muy general que no se restringe a la medida de una variable y a su posterior envío a un equipo receptor. Al contrario, los sistemas multivariable son muy frecuentes y pueden implicar pasos adicionales de multiplexado de señales y la presencia de bloques más o menos inteligentes a ambos extremos del mero enlace de transmisión según se indica en la figura 20.16.

A pesar de que el sistema de la figura 20.16 es un caso general en el que fácilmente pueden enmarcarse muchos sistemas de instrumentación, también hay casos más sencillos que sólo requieren algunos bloques. Así, para un buen número de casos, no será necesaria ningún tipo de modulación, en otros se tratará de un sistema monovariable, etc. Para ir barriendo las diversas posibilidades, comenzaremos en este capítulo por analizar el más sencillo de todos desde el punto de vista del medio de transmisión e iremos avanzando en la siguiente hasta sistemas más complejos.

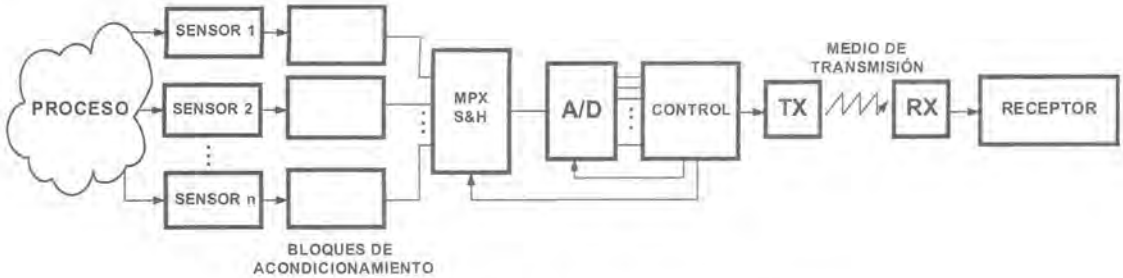


Figura 20.16. Sistema de telemedida multivariable.

20.3 Bucles de tensión y bucles de corriente

Después del tratamiento de la señal producida por cualquier sensor se suele obtener un valor de tensión que debería ser proporcional a la variable a medir y que tendrá un nivel suficiente. Esta señal se puede emplear directamente para su representación en un *display* o en cualquier otro tipo de indicador o bien ser enviada hasta otro lugar para que otro equipo la procese.

20.3.1 BUCLES DE TENSIÓN

La primera tentación para enviar esa señal que es, en sí misma, la información de la variable medida es la conexión directa del circuito de tratamiento al equipo que la procesará (figura 20.17a).

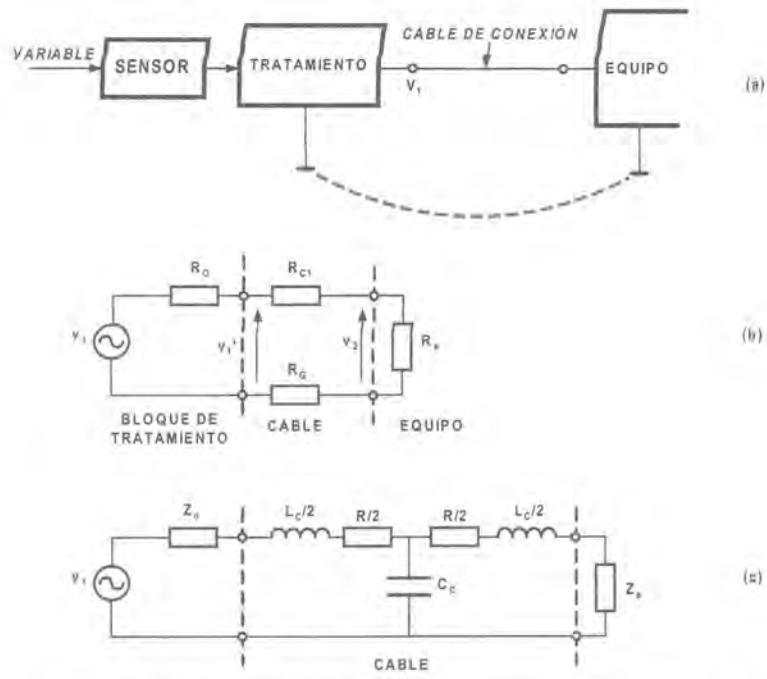


Figura 20.17. Bucles de tensión: (a) conexión; (b) equivalente básico sólo resistivo; (c) equivalente completo.

Si así lo hacemos nos podemos encontrar con algunas sorpresas: en primer lugar, si consideramos el equivalente Thévenin de la salida del bloque de tratamiento, la resistencia de entrada del equipo que recibe la señal y que los cables no son ideales, sino que presentan un valor de resistencia por unidad de longitud, nos encontraremos con un circuito "real" como el de la figura 20.17b. En él, la tensión que recibe el equipo de llegada no es la ideal v_1, sino que es v_2:

$$v_2 = \frac{R_e}{R_e + R_C + R_G + R_O} v_1 < v_1 \tag{20.3}$$

donde R_C es la resistencia del cable, R_G es la resistencia del retorno por tierra de la señal y R_O es la resistencia de salida del circuito de tratamiento de la información. Esto supone que el equipo receptor "ve" una señal menor de la que debería llegarle en condiciones ideales lo que supone la presencia de un error que se puede cuantificar en términos relativos como:

$$\varepsilon = \frac{R_C + R_G + R_O}{R_e + R_C + R_G + R_O} \tag{20.4}$$

y que crece en la medida en que la resistencia de entrada R_e sea pequeña o que sean grandes la resistencias del cable, del retorno por tierra y de salida. Como quiera que R_C es un parámetro que crece linealmente con la distancia, el problema se agrava a medida que el equipo receptor está más lejos del emisor. En cualquier caso, incrementar la impedancia de entrada del bloque receptor, reducirá el error producido.

Pero ahí no terminan los problemas ya que, en realidad, los cables llevan más parásitos asociados además de los estrictamente resistivos y que se pueden representar como inductancias y capacidades (figura 20.17c). Incluso, las resistencias de salida y entrada son, en realidad, impedancias lo que aún complica más

el problema ya que aporta caídas de tensión adicionales en alterna y limita el ancho de banda de las señales que pueden circular por el sistema. Como quiera que los diversos parásitos del cable crecen linealmente con la longitud, todos estos efectos se acrecientan a medida que aumenta la distancia que separa el emisor del receptor, haciendo que el error sea mayor.

A pesar de todo lo dicho, en distancias cortas o para conectar tarjetas dentro de un mismo equipo, se utiliza con éxito este medio que consiste en enviar la información utilizando un cable de cobre como medio de transmisión mediante una señal de tensión y que se denomina *bucle de tensión*.

Muchos sistemas y equipos producen salidas en tensión que otros equipos o sistemas tendrán que leer; para facilitar la interconexión de unos y otros se han establecido una serie de valores estándar de tensiones para representar los valores de las variables, normalmente definidos por las tensiones extremas que representan, respectivamente, los valores mínimo y máximo de la variable correspondiente. Así, tenemos bucles de tensión de:

a) 0 a 5 V.
b) –5 a 5 V.
c) 0 a 10 V.
d) –10 a 10 V.
e) 0 a 24 V.
f) –2,5 a 2,5 V.

Es posible encontrar sistemas que admitan una o varias de estas escalas aunque alguna de ellas es muy habitual como la de 0 a 24 V que está muy extendida a nivel industrial.

Sin embargo, el principal problema de los bucles de tensión radica en la posibilidad de introducción de interferencias externas en el propio bucle. Las interferencias externas se considerarán como ruidos introducidos en el sistema y que se representarán como una fuente de tensión en serie v_n con el propio lazo y como una fuente de corriente i_n en paralelo con él (esto se demostrará en el último tema teórico de este libro) según se indica en la figura 20.18 en la que se ha usado sólo el modelo de resistencias y se ha considerado toda la resistencia del cable y del retorno como R_C.

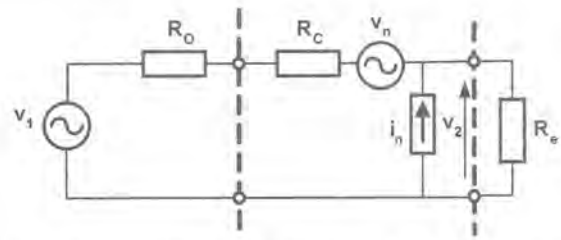


Figura 20.18. Modelo de transmisión de señal en tensión incluyendo las fuentes de ruido.

La tensión que se desarrolla sobre la resistencia de entrada del receptor R_e será:

$$v_2 = \frac{R_e}{R_e + R_O + R_C}(v_1 + v_n) + \frac{R_e(R_O + R_C)}{R_e + R_O + R_C} i_n \qquad (20.5)$$

La relación S/N será, partiendo de la expresión anterior, la siguiente:

$$S/N = 20\log \frac{v_1}{v_n + (R_O + R_C) i_n} \qquad (20.6)$$

Recordando que la resistencia del cable viene dada por:

$$R_C = \rho \frac{2l}{S}$$

donde l es la longitud del enlace (2l representa el cable más el retorno), S es la sección y ρ, la resistividad, la expresión (20.6) se puede escribir como:

$$S/N = 20\log \frac{v_l}{v_n + (R_O + 2\rho l/S)i_n} \qquad (20.7)$$

Del análisis de la expresión (20.6) se deduce una serie de importantes consideraciones:

a) La impedancia de entrada del circuito receptor no afecta a la relación señal-ruido por lo que deberá usarse un valor elevado para garantizar que el error introducido por la expresión (20.4) se vea reducida. Esto, si bien resulta cierto según la expresión (20.6), no lo es del todo pues hay que recordar que cualquier resistencia introduce un nivel de ruido térmico que crece con la raiz cuadrada de su valor óhmico.

b) La resistencia de salida del circuito emisor debe ser lo más reducida que sea posible para conseguir reducir el efecto de la corriente de ruido i_n.

c) La resistencia del cable debe ser, como la anterior, tan pequeña como sea posible lo que se traduce en **cables cortos** y de **sección grande**. Esto parece un contrasentido puesto que el uso de un bucle de tensión con R_e grande supone una circulación débil de corriente; ¿para qué, entonces, una sección elevada? La respuesta es muy simple: para que sobre ella no se desarrollen caídas de tensión debidas a la corriente de ruido.

Como consecuencia de todo esto, los enlaces mediante bucles de tensión sólo sirven para distancias cortas.

EJERCICIO 20.2

Determine el valor de la relación S/N cuando se pretende enviar una señal de tensión 0÷5 V desde un punto a otro en función de la distancia, considerando que la resistencia de salida del circuito emisor puede considerarse nula utilizando cables de cobre de 0,1 mm² de sección sabiendo que la resistividad del cobre a la temperatura de trabajo es de 0,018 Ωmm²/m. ¿Qué ocurriría si el cable empleado tuviese 1 mm² de sección? Supóngase que el valor de la tensión de ruido inducida es de 1 μV y el de la corriente 1 μA, independientes de la longitud del enlace[5].

SOLUCIÓN:

Como es habitual, vamos a hacer el cálculo en el fondo de la escala y consideremos primero el caso de la menor sección; aplicando la expresión (20.7) se obtiene que el valor de la relación S/N para este caso es:

$$S/N = 20\log \frac{5}{10^{-6} + 0{,}018 \cdot 2 \cdot l/0{,}1 \cdot 10^{-6}} = 20\log \frac{5}{10^{-6} + 3{,}6 \cdot 10^{-7} l}$$

En la figura 20.19 se representa la anterior función, observándose una fuerte degradación de la relación S/N.

[5] Esto no es demasiado correcto y, en general, las tensiones y corrientes de ruido se incrementarán a medida que crece la distancia del enlace; en este caso, hemos hecho esta suposición para simplificar el problema y destacar el efecto del comportamiento del propio bucle.

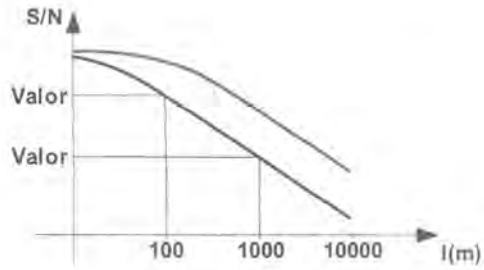


Figura 20.19

Si se hubiese utilizado el cable de mayor sección, la expresión de la relación S/N sería un poco mejor:

$$S/N = 20 \log \frac{5}{10^{-6} + 0{,}018 \cdot 2 \cdot 1 \cdot 10^{-6}} = 20 \log \frac{5}{10^{-6} + 3{,}6 \cdot 10^{-8}\, l}$$

expresión que se representa también en la figura 20.19, demostrando que es mejor utilizar cables de la mayor sección que sea posible aunque, claro está, esto redunda negativamente en el coste de la instalación.

El modelo de ruido utilizado hasta aquí está muy simplificado puesto que, para un análisis más completo, debería considerarse el modelo del cable que incluye los parásitos capacitivos e inductivos distribuidos con fuentes de corriente y tensión de ruido también distribuidas. No obstante, dado lo impredecible e incontrolable del ruido, resultaría muy difícil conocer con alguna exactitud el efecto final incluso con ese modelo. Por tanto, el estudio realizado puede ser suficiente para conocer cualitativamente los efectos sobre la relación S/N y establecer algún tipo de limitaciones al uso de los bucles de tensión. Dejamos a la curiosidad del lector la realización de un estudio similar al anterior pero en el que las fuentes de ruido incrementen su nivel de forma proporcional a la longitud del enlace.

En cualquier caso, para la implementación de bucles de tensión se usarán pares trenzados para conseguir reducir el efecto de las posibles fuentes externas de ruido; son habituales los cables de tipo STP o FTP excepto para distancias muy cortas (como el caso de cableado entre tarjetas de un mismo equipo) en que se pueden emplear cables de tipo UTP.

20.3.2 BUCLES DE CORRIENTE

La otra posibilidad para transmitir la información por un cable de cobre es hacerlo mediante una señal de corriente que funciona de forma similar al bucle de tensión pero asignando a cada valor de la información un nivel de corriente comprendido entre dos valores extremos determinados; el cable de cobre hará llegar la corriente hasta el extremo receptor en que desarrollará una tensión sobre la resistencia de llegada. Un sistema de este tipo se denomina *bucle de corriente*; se muestra un esquema básico en la figura 20.20a.

La tensión desarrollada sobre la resistencia de llegada será:

$$v_2 = \frac{R_O R_e}{R_O + R_e} i_p \qquad (20.8)$$

Si se considera la resistencia del cable, R_C englobando al camino de ida y al de retorno según se indica en la figura 20.20b, la expresión (20.8) se transforma en:

$$v_2 = \frac{R_O R_e}{R_O + R_C + R_e} i_p \qquad (20.9)$$

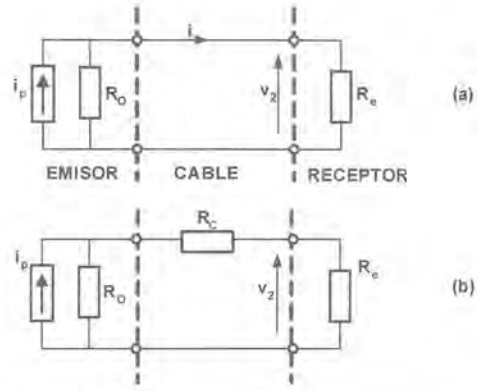


Figura 20.20. Bucle de corriente (a) y modelo incluyendo la resistencia del cable (b).

Esta expresión introduce un error en la tensión que se recibe sobre R_e muy similar, en principio, al error producido en el caso del bucle de tensión representado por las expresiones (20.3) y (20.4). Sin embargo, esta similitud es sólo aparente puesto que en aquel caso una buena fuente de tensión en la emisión ($R_O = 0$) no eliminaba el error mientras que, en este caso, una buena fuente de corriente ($R_O = \infty$) hace que v_2 sea el valor ideal, es decir, $R_e i_p$ con lo que el error debido a la resistencia del cable se hace nulo.

Igual que en el caso de los bucles de tensión, hay valores estándar para los bucles de corriente que se indican mediante los valores de los extremos que representarán los límites inferior y superior de la variable a transmitir. Son los siguientes:

a) –8 a 12 mA.
b) 0 a 20 mA.
c) 4 a 20 mA.

También han existido algunos más como los 0÷60 mA en algunas aplicaciones militares pero básicamente son estos tres los más habituales aunque, por concretar más aún, los dos últimos. La mayoría de las señales de instrumentación que se transmiten por un cable de cobre se envían mediante bucles de corriente de 0÷20 mA o 4÷20 mA. Incluso, su uso se ha hecho extensivo al envío de señales digitales en muchas aplicaciones industriales.

En la recepción se suelen utilizar resistencias de bajo valor, de 250 ó 500 Ω, sobre las que los bucles de corriente producen los niveles de tensión que se indican en la tabla 20.3 y que, en el caso, de usar 0÷20 mA, se corresponden directamente con valores normalizados de niveles de tensión.

Bucle de corriente	R_e	Niveles de tensión
0÷20 mA	250 Ω	0÷5 V
	500 Ω	0÷10 V
4÷20 mA	250 Ω	1÷5 V
	500 Ω	1÷10 V

Tabla 20.3. Conversiones de corriente estándar a tensión estándar con las resistencias habituales.

Aunque ya hemos visto una ventaja del bucle de corriente respecto al de tensión, podríamos pensar que empleando en ambos un buen cable y una buena fuente de emisión quedan más o menos iguales o que, en

cualquier caso, el error producido en el de tensión es sistemático y podría compensarse sin más que incrementando un poco la ganancia en cualquier punto de la cadena. Para completar la valoración, veamos qué pasa con la relación S/N.

En la figura 20.21 se considera el modelo de ruido que se empleó para el bucle de tensión como una fuente de tensión en serie y una fuente de corriente en paralelo.

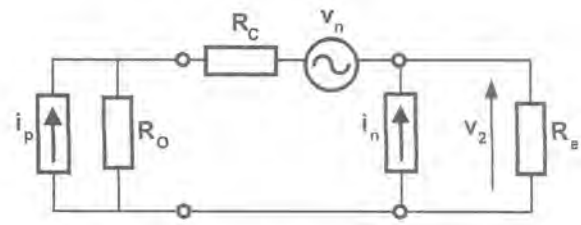


Figura 20.21. Bucle de corriente incluyendo el modelo de ruido.

La tensión producida sobre la resistencia de llegada será:

$$v_2 = \left(\frac{R_O}{R_O + R_C + R_e} i_1 + \frac{R_O + R_C}{R_O + R_C + R_e} i_n \right) R_e + \frac{R_e}{R_O + R_C + R_e} v_n \qquad (20.10)$$

Con ello, la relación S/N resultará ser:

$$S/N = 20 \log \frac{R_O i_1}{(R_O + R_C) i_n + v_n} \qquad (20.11)$$

Esta expresión se parece bastante a la que se obtenía para el bucle de tensión excepto en un aspecto muy importante: si la fuente de corriente es ideal ($R_O = \infty$), la relación S/N no dependerá de la resistencia del cable, es decir, no dependerá de su longitud y, además, anulará el efecto de la fuente de tensión de ruido quedando reducida la relación S/N al cociente de la corriente de señal y la de ruido. Como quiera que puede conseguirse sin demasiadas dificultades una fuente de corriente bastante buena, **el bucle de corriente permite obtener mejores relaciones S/N que el de tensión**.

EJERCICIO 20.3

Determine el valor de la relación S/N cuando se pretende enviar una señal de corriente 0÷20 mA desde un punto a otro en función de la distancia, considerando que la resistencia de salida del circuito emisor puede considerarse nula, utilizando cables de cobre de 0,1 mm² de sección sabiendo que la resistividad del cobre a la temperatura de trabajo es de 0,018 Ωmm²/m. ¿Qué ocurriría si el cable empleado tuviese 1 mm² de sección? Supóngase que el valor de la tensión de ruido inducida es de 1 μV y el de la corriente, 1 μA, independientes de la longitud del enlace.

SOLUCIÓN:

Obsérvese que el ejemplo es muy similar al 20.2 excepto en que ahora se está empleando un bucle de corriente en lugar de uno de tensión. Como quiera que consideramos ideal la fuente de corriente de emisión, la relación S/N determinada por la expresión (20.11) quedará reducida a:

$$S/N = 20\log\frac{i_l}{i_n} \Rightarrow S/N = 20\log\frac{20\cdot10^{-3}}{10^{-6}} = 86\ dB$$

que resulta independiente de la distancia del enlace. Es interesante comparar el resultado de este ejemplo con las gráficas obtenidas en el ejercicio 20.2 para comprobar cómo se comportan los dos sistemas de transmisión desde el punto de vista de la relación S/N. También debemos recordar que hemos efectuado una simplificación un tanto grosera, asumiendo que la fuente de ruido es independiente de la longitud del enlace lo que sabemos que es falso; aun así, el bucle de corriente "elimina" el efecto de la fuente de tensión de ruido y limita el efecto de la fuente de corriente de ruido.

Sin embargo, el bucle de corriente no está exento de problemas; uno de ellos se presenta con la resistencia de llegada que "lee" la corriente que circula por el bucle. El valor de la tensión que se reciba depende de su valor por lo que este debe resultar lo más estable que sea posible para reducir la incertidumbre. Así, habrá que tener en cuenta el valor de la tolerancia y las posibles variaciones de su valor debido a cambios de temperatura inducidos por el ambiente o por efecto Joule.

EJERCICIO 20.4

Un sistema que recibe señal de un bucle de corriente 0÷20 mA posee una resistencia de entrada de 500 Ω (@ 25°C) cuyo valor tiene una tolerancia de ±1% y posee un coeficiente térmico tal que provoca un incremento de su valor de 0,100 Ω cada grado centígrado. Determine el máximo error a fondo de escala que se producirá en la recepción admitiendo variaciones de la temperatura ambiente de 25 a 45 °C y despreciando el autocalentamiento por efecto Joule.

SOLUCIÓN:

El peor caso, es decir, cuando se producirá el máximo error, corresponde a la situación más extrema en que se acumulen los efectos del incremento de la temperatura y de máxima tolerancia. Para calcular el máximo error consideraremos las condiciones óptimas y las extremas.

En condiciones de 25°C de temperatura y con el valor exacto de resistencia, la corriente de 20,0 mA ocasionará una tensión de salida de 10,0 V. Por otro lado, la máxima resistencia que se podrá tener corresponde al caso en que la tolerancia sea +1%, es decir, 505 Ω y que se encuentre a la máxima temperatura (45°C). Para ese caso, el valor de resistencia se habrá incrementado en 2,00 Ω, resultando un valor total de resistencia de 507 Ω; la corriente de 20,0 mA causará una tensión de 10,1 V por lo que el error introducido será del 1% aproximadamente.

En general, para disminuir los errores debidos a la resistencia de llegada, se procurará utilizar resistencias estables térmicamente y de alta precisión (tolerancias de 0,5 ó 0,1%); a veces, puede ser recomendable el uso de parejas de resistencias en serie o en paralelo que posean coeficientes térmicos inversos, uno negativo y otro positivo para que se compensen parcialmente los efectos del cambio de la temperatura. Finalmente, se usará un dispositivo capaz de disipar la potencia desarrollada que, en el caso de la máxima corriente sobre una resistencia de 500 Ω llegará a ser de 0,2 W. A este respecto, no sólo deberá usarse una resistencia capaz de disiparla, sino que convendrá usar una de bastante más potencia para que su resistencia térmica sea menor y, en consecuencia, disipe la potencia sin provocar un excesivo autocalentamiento que pueda modificar su valor óhmico con la introducción del consiguiente error.

Para la transmisión de señales por bucle de corriente se suele utilizar par trenzado de cualquiera de los tipos que hemos visto, aunque se prefiere utilizar alguno que tenga pantalla como el de tipo STP o el FTP ya que permite reducir los acoplamientos capacitivos que dan lugar a corrientes de ruido (véase el capítulo 28). Así, la representación de un bucle de corriente o lazo de corriente, es la que se muestra en la figura 20.22.

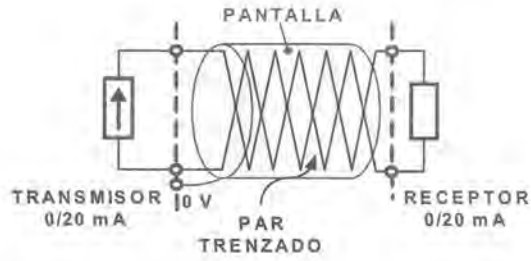


Figura 20.22. Bucle de corriente que emplea un par trenzado y apantallado.

20.3.3 CONVERSIÓN V/I

La utilidad de los bucles de corriente como sistemas para enviar señales de un punto a otro ha quedado ya establecida; por otro lado, tenemos que la mayoría de las señales en un sistema electrónico son de tensión por lo que suele ser necesario efectuar una conversión entre una y otra magnitud. Veremos ahora cómo se puede efectuar este paso.

En una primera aproximación, la *conversión tensión-corriente* (V/I) se podría hacer con circuitos como los que se muestran en la figura 20.23.

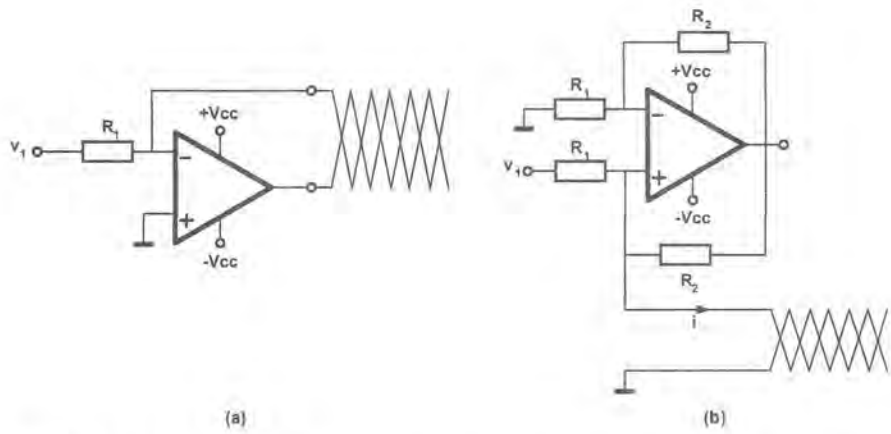


Figura 20.23. Circuitos de conversión V/I: (a) flotante; (b) referido a masa.

En cualquiera de los dos casos de la figura, la corriente que circula por la salida del sistema (por el bucle de corriente) resulta ser v_1/R_1 con lo que la conversión es muy cómoda. Sin embargo, el circuito de la figura 20.23a presenta un posible problema como es el que ninguno de sus dos extremos de salida sean masa; esta salida flotante podría ser problemática a la hora de conectar la resistencia en la llegada del bucle que no podría tener ninguno de sus terminales conectados a masa, so pena de cortocircuitar la salida del operacio-

nal. El segundo circuito es mucho más cómodo pero obliga a utilizar pares de resistencia iguales lo que, en la práctica, puede ser bastante complejo.

Como problema adicional de ambos circuitos está el hecho de la limitación de la corriente máxima que proporciona el operacional y que podría obligar a utilizar amplificadores operacionales de más potencia o a añadir etapas de transistores en colector común a la salida del operacional para poder proporcionar los valores necesarios para el bucle de corriente.

Un problema común a ambos circuitos y, en general, a cualquier convertidor V/I radica en el máximo valor de la carga que se puede conectar a la salida. Recordemos que los operacionales no pueden proporcionar tensiones que rebasen los extremos de alimentación y lo más común es que sólo puedan acercarse a uno o dos voltios de distancia. Por lo tanto, la conversión V/I será efectiva siempre y cuando la tensión en la carga no obligue al operacional a superar su máxima excursión de tensión de salida. A este efecto, es necesario contabilizar la propia resistencia de carga del circuito de recepción y la resistencia del cable de transmisión.

EJERCICIO 20.5

Sea el circuito de la figura 20.23b en el que se ha utilizado un operacional alimentado a ±12 V capaz de producir en su salida tensiones desde –10 V hasta +10,5 V. Suponiendo que $R_1 = 250\ \Omega$, $R_2 = 100\ \Omega$, que la tensión de entrada v_I se encuentra comprendida entre 0 y 5 V, determinar el máximo valor de la carga del bucle de corriente de salida. Si la resistencia de llegada es de 250 Ω y se utiliza un cable de cobre de 0,1 mm^2 de sección con $\rho = 0{,}018\ mm^2/m$, ¿cuál será la máxima distancia que puede alcanzar el bucle de corriente sin que se sature el operacional? ¿Cómo podría incrementar el alcance?

SOLUCIÓN:

El caso más desfavorable será el que corresponde a la máxima corriente circulando por el bucle que corresponde a la situación en que $v_I = 5$ V; en ese caso, $i = 5/250 = 20$ mA. Si llamamos v_+ a la tensión en la entrada no inversora del operacional, tenemos que $v_+ = i \times R_C$ considerando que R_C es la resistencia de carga que supone el bucle de corriente (figura 20.24).

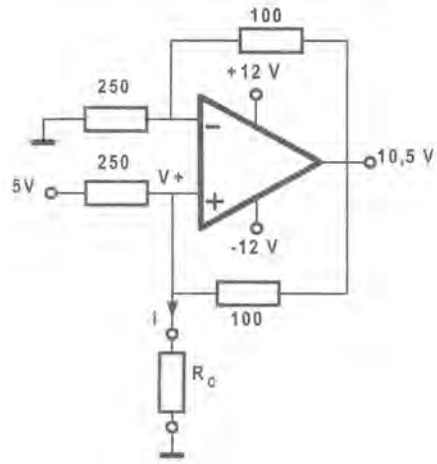


Figura 20.24

Ahora nos ponemos en el extremo de tensión de salida del operacional, es decir, 10,5 V. Como no circula corriente por la entrada no inversora del operacional, tenemos que:

$$\frac{5-v_+}{250}+\frac{10{,}5-v_+}{100}=20\ \text{mA}$$

Despejando el valor de v_+ se obtiene 7,5 V lo que significa que la máxima carga que admite el sistema será de 7,5 V/20 mA = 375 Ω. Como quiera que la resistencia de carga es de 250 Ω, sólo quedan disponibles para el cable 125 Ω, que corresponde a una longitud de enlace dada por:

$$125=0{,}018\frac{2l}{0{,}1}\Rightarrow l=350\ \text{m}$$

lo que es una distancia de enlace apreciable; si se pretendiese aumentarla, habría que actuar en alguno de los siguientes sentidos:

a) Incrementar la sección del hilo de cobre para reducir la resistencia efectiva del cable y, así, descargar el bucle de corriente.

b) Modificar las resistencias de realimentación R_2 del operacional hacia menores valores.

c) Disminuir el valor de la resistencia de llegada aunque eso nos impidiese obtener una conversión a tensión a valores estandarizados muy simple.

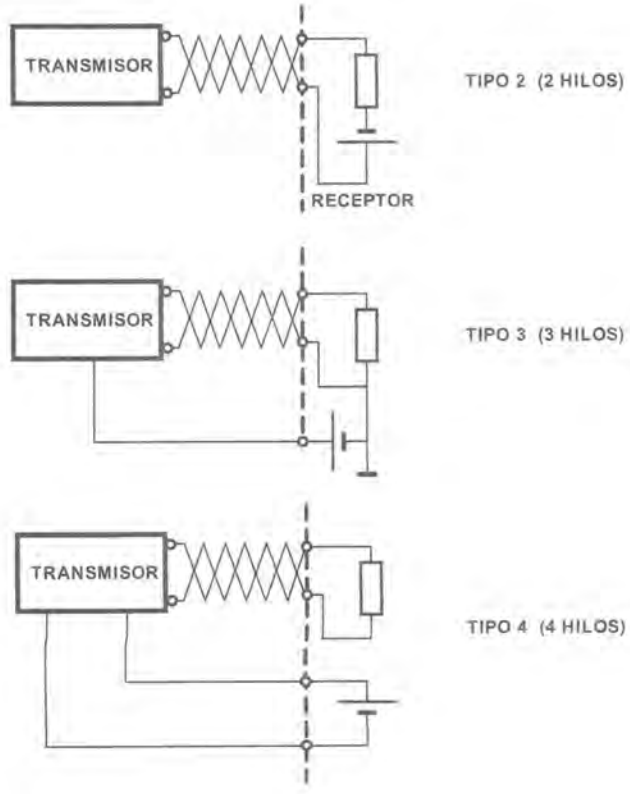


Figura 20.25. Diferentes tipos de alimentación en los bucles de corriente.

La capacidad de un bucle de corriente de proporcionar una tensión alta en su salida permite cargar más o menos el circuito. Desde este punto de vista se suelen dividir en tres clases L, H y U que quedan definidas

por la tensión de alimentación y por la carga que es posible situar; módulos de la misma clase son intercambiables entre sí [20.8].

En el caso anterior hemos supuesto una alimentación tradicional pero, en muchos casos, el sistema que envía la señal es un sensor que incluye los circuitos de adaptación y el transmisor del bucle de corriente (*sensor-transmisor* o, simplemente, *transmisor*). En estos casos, no suele ser normal que se incluya una alimentación específica para el sensor, sino que este se alimentará de la misma fuente que el resto del sistema; para simplificar el cableado se puede alimentar de forma que el número de hilos necesario se reduzca, disminuyendo los costes. En la figura 20.25 se muestran los tres tipos de transmisores (tipos 2, 3 y 4) [20.3], [20.8] según cómo se conecte la fuente de alimentación.

20.3.4 CONVERSIÓN I/V

A la llegada del bucle de corriente hay que recuperar la tensión como magnitud puesto que esta es mucho más cómoda para poder ser utilizada. Para ello, se recurre al convertidor I/V por excelencia que es la resistencia. En la figura 20.26 se muestra el circuito básico de conversión formado por una resistencia del valor óhmico adecuado (250 ó 500 Ω) y un amplificador operacional como seguidor de emisor que permite la carga del circuito sin que perturbe al valor recibido.

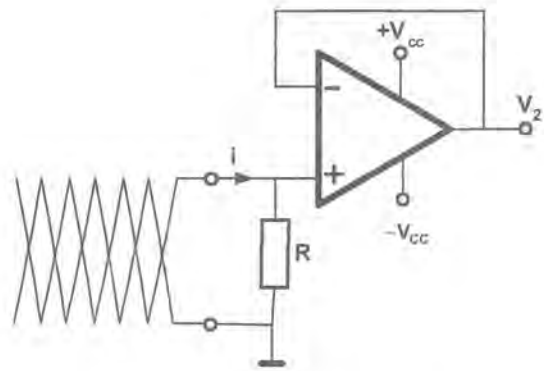


Figura 20.26. Circuito de llegada de un bucle de corriente.

El valor de la tensión de salida será:

$$v_2 = Ri \tag{20.12}$$

hemos dicho que los bucles de corriente más habituales son los de 0÷20 mA y 4÷20 mA y, aunque en el primer caso está claro que la conversión a tensión es directa, en el segundo podemos preguntarnos: ¿por qué ese valor de 4 mA? ¿Aporta algo? Lo cierto es que sí y que no vamos a complicarnos la vida para no obtener ninguna ventaja.

Podemos pensar en un sistema receptor al que está llegando la información mediante un bucle de corriente 4÷20 mA. En cualquier punto de su margen de trabajo, siempre llega algo de corriente y se obtiene, consecuentemente, algo de tensión en la llegada. Si, por cualquier circunstancia se rompiese la línea, llegaría un valor de 0 mA y no produciría tensión con lo que el receptor detectaría el problema y daría la correspondiente señal de alarma. Esto no ocurriría en el caso de un bucle 0÷20 mA en que el 0 es un valor más del

margen de trabajo del sistema y nunca sabríamos si está llegando un cero porque el emisor lo envía o porque la línea se ha roto.

De esta manera, el sistema 4÷20 mA denominado bucle de corriente con *cero vivo*, permite la detección de la *línea rota*[6]. En la figura 20.27 se muestra un circuito de ejemplo para la recepción de bucle de corriente 4÷20 mA y conversión a 1÷5 V, con detección de línea rota mediante un comparador que produce la correspondiente señal de alarma.

Lógicamente, el uso del bucle de corriente 4÷20 mA reduce el margen de trabajo en un 20% y, si la cuantificación del valor se realiza en el margen 0÷5 V, se estará perdiendo una parte de la escala, dedicándola a supervisión del enlace; para evitar este problema que haría perder resolución a los sistemas de conversión A/D, podemos incluir un sencillo circuito analógico que regenere una señal de 0 a 5 V.

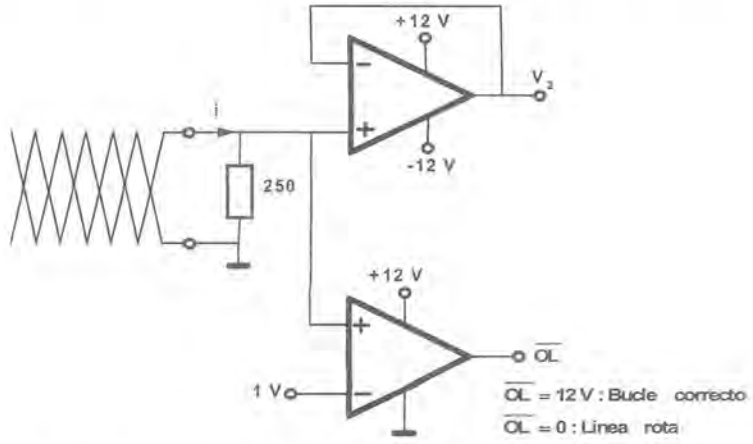


Figura 20.27. Recepción de un bucle de corriente con detección de línea rota.

20.3.5 CONSIDERACIONES PRÁCTICAS SOBRE LOS BUCLES DE CORRIENTE

Cuando se pretende utilizar un bucle de corriente en un sistema de instrumentación, no suele ser normal el diseño del sistema de emisión tal y como se ha detallado en el apartado anterior, sino que lo más habitual es utilizar alguno de los bloques que, a este fin, han desarrollado casi todos los fabricantes de circuitos integrados analógicos. Así, *Analog Devices*, *Texas Instruments*, *National Semiconductor* y otros más, disponen de circuitos de conversión V/I con entradas de tensiones estándar y salidas en cualquiera de los dos bucles de corriente más típicos. En el CD que acompaña al libro se pueden ver algunos ejemplos de este tipo de circuitos.

También hay otro tipo de conversores V/I e I/V disponibles, pero no implementados como circuitos impresos sino como bloques con entradas y salidas estándar y previstos para sujeción en carrileras normalizadas. En este caso no cabe hablar de componentes electrónicos sino de pequeños equipos que forman parte de sistemas de instrumentación más complejos o de control.

Un caso muy especial pero de indudable interés de bucle de corriente es el sistema de bucle de corriente aislado que permite enviar informaciones entre equipos con masas no comunes, ya sea porque los potencia-

[6] Aunque la rotura física de la línea no suele ser muy habitual, sí que se puede producir un desconexionado de algún punto del bucle o, quizá una salida del margen de tensiones válido.

les de referencia respectivos estén muy alejados o porque se pretende romper una línea de circulación de ruido. En estos casos, se recurre al empleo de bucles aislados en los que no hay contacto galvánico entre emisor y receptor. Todos los sistemas de este estilo no son muy diferentes de los presentados en el apartado 20.3.2 excepto en el uso de amplificadores aislados como los vistos en el capítulo 3.

A la hora de implementar un bucle de corriente aislado hay que tener presente de dónde se va a alimentar cada parte del sistema. En la figura 20.28 se muestra un emisor y un receptor de bucle de corriente aislados [20.1], [20.4].

El caso más complejo supone la presencia de tres fuentes de alimentación diferentes y la presencia de dos puntos de aislamiento. En estas condiciones, el sistema adquiere la máxima versatilidad puesto que permite la utilización de un emisor o un receptor no aislado sin romper el aislamiento entrada/salida general. Sin embargo, un sistema como este implica la presencia de tres fuentes de alimentación y de una conexión de tipo 4 [20.8].

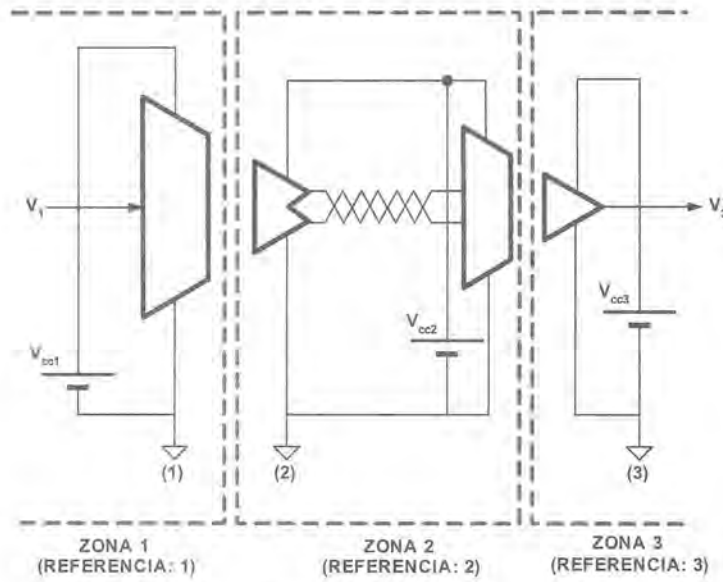


Figura 20.28. Bucle de corriente doblemente aislado.

Una simplificación del sistema anterior se muestra en la figura 20.29 en la que la presencia de aislamiento se limita a uno de los extremos.

La solución de la figura 20.29a sólo precisa dos fuentes de alimentación aunque la conexión precisa cuatro hilos (como está dibujada) o tres hilos si la resistencia de llegada estuviera referida a la masa 1; este caso es el típico en el que el sensor transmisor se alimenta desde el sistema al que tiene que enviar la información. En la figura 20.29b, hay una alimentación en cada lado y el circuito de llegada del lado receptor se alimenta desde la misma fuente que el transmisor; nuevamente se ha dibujado la solución a cuatro hilos pero es factible emplear una solución a tres hilos si la resistencia de recepción está conectada en uno de sus extremos a la masa 1.

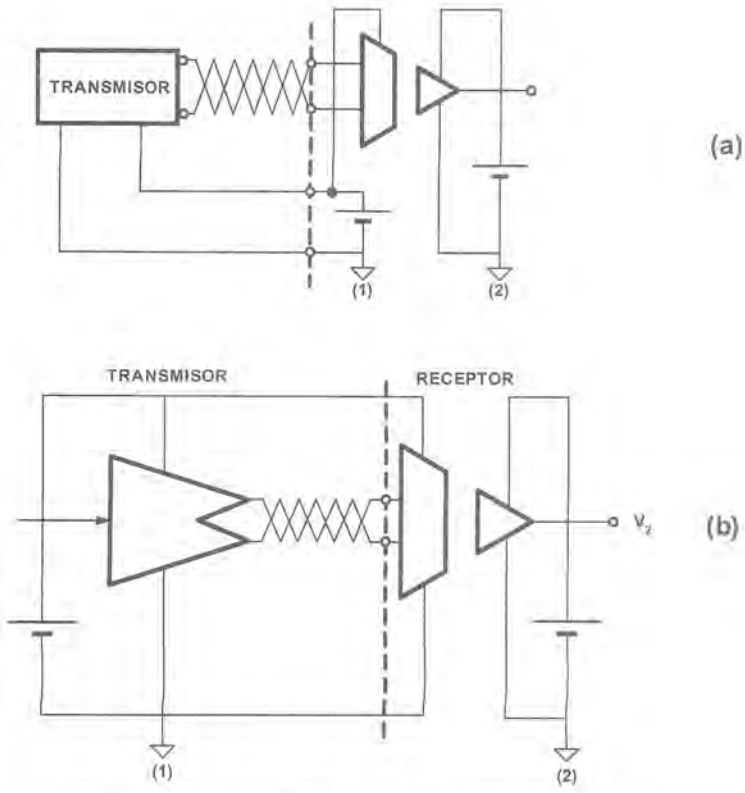


Figura 20.29. Bucles de corriente aislados sencillos.

Resumen

Este capítulo presenta un breve recorrido por los diversos métodos de transmitir la información desde un origen a un destino y ha introducido el concepto general de telemedida que no es más que la aplicación de los sistemas de transmisión a la instrumentación electrónica para conseguir medidas remotas.

Muy someramente se han tocado diversos medios de transmisión, tanto los que emplean un soporte físico (cable de cobre, fibra óptica) como los que no lo precisan (radio, luz). Aunque un estudio pormenorizado de la mayoría de estos medios se escapa por completo a los objetivos del presente libro, sí que se han hecho algunos comentarios que permiten tomar decisiones acerca de cuál emplear en cada caso.

Dentro de los sistemas guiados —los que precisan un soporte físico— se ha comentado la necesidad del trenzado y del apantallado de los cables de cobre, las ventajas inherentes a la fibra y las limitaciones e inconvenientes de su uso. Los sistemas no guiados constituyen un caso más complejo que los anteriores la mayoría de las veces por lo que no se ha entrado en detalles.

Se ha comentado brevemente el interés de la codificación de la información, la complejidad que se añade en el diseño del sistema y las ventajas que aporta desde el punto de vista de la seguridad.

Para concluir el capítulo —y dado el elevado uso industrial— se ha tratado con detalle el empleo de bucles de corriente como una alternativa para resolver algunos serios problemas ocasionados por los bucles de

tensión. En este caso, se han comentado los valores estadarizados más típicos, los circuitos que permiten las conversiones entre tensión y corriente, sus limitaciones y consideraciones de diseño. Se ha finalizado con una serie de aspectos relacionados directamente con los bucles de corriente desde el punto de vista de su aplicación directa a los sistemas de instrumentación.

Problemas propuestos

20.1 Empleando la fibra óptica y las condiciones del ejercicio 20.1, indique la solución para realizar un enlace de 6 km.

20.2 En el caso del problema anterior, ¿cuál debería ser la atenuación máxima de la fibra para conseguir que el enlace pueda alcanzar los 10 km?

20.3 La salida de un sistema proporciona una tensión entre 0 y 12 V para codificar la información de un sensor y se quiere llevar a un sistema de procesamiento que está alejado del primero una distancia de 20 m (de enlace de cable). Si el cable empleado tiene una resistividad de 0,18 $\Omega mm^2/m$ y la sección que se emplea es de 0,1 mm^2, determine cuál es la tensión que recibe el sistema de procesamiento si su impedancia de entrada fuese de 10 kΩ. ¿Y si fuese de 1 MΩ? Evalúe la atenuación en dB.

20.4 En las condiciones del problema anterior se supone que el enlace está sometido a unas condiciones de ruido dentro de su ancho de banda de 10 nA y 10 μV. ¿Cuál es la relación S/N del enlace?

20.5 ¿Cuál es la distancia máxima de enlace que permite una relación S/N mejor de 50 dB para las condiciones del problema anterior?

20.6 Si duplicamos la distancia máxima de enlace del problema 20.5, ¿cuál es la relación S/N obtenida? ¿Cómo se podría mantener la relación S/N en 50 dB? Sugerencia: estudie las posibilidades de subir la tensión de salida y mejorar el cable.

20.7 Considere la posibilidad de emplear un bucle de corriente 4/20 mA con 500 Ω en la resistencia de llegada para ejecutar el enlace del problema 20.3. Compare los resultados.

20.8 Determine la relación S/N en un bucle de corriente 4/20 mA con una resistencia de llegada de 250 Ω en el que aparece una corriente total de ruido de 1 nA y una tensión de ruido de 1 μV en el total del ancho de banda del sistema.

20.9 Diseñe un sistema de detección de línea rota para el bucle de corriente del problema anterior.

20.10 Determine el efecto de la temperatura sobre la tensión que recibe la resistencia de llegada de 250 Ω de un bucle de corriente si se trata de una resistencia metálica que sigue la ley $R(T) = R_0(1 + \alpha T)$. Determine la función que relaciona el error con la temperatura. ¿Cuál es el parámetro que más le afecta?

20.11 Diseñe un convertidor V/I para transformar una señal de –5 a 5 V en otra de 4/20 mA empleando un amplificador operacional de tipo 741 (ver características en el CD adjunto).

20.12 ¿Qué precauciones se deben tener en cuenta para el óptimo diseño del sistema del problema anterior?

21

Modulación y demodulación

21.1 Modulación con portadora analógica y moduladora analógica
21.2 Modulación con portadora analógica y moduladora digital
21.3 Modulaciones con portadora digital y moduladora analógica

La modulación y demodulación de señales está presente, de algún modo, en el diseño de la mayor parte de los sistemas de instrumentación. Supóngase, por ejemplo, que se desea realizar un detector óptico de barrera. En principio se podría utilizar una fuente de luz continua (por ejemplo, un LED alimentado en continua) enfrentada con un sensor de luz (por ejemplo, un fotodiodo) de forma que se sabría que un objeto está presente cuando se interrumpe el haz de luz. El objeto modifica, en definitiva, la amplitud de la luz que llega al fotodiodo. A esta acción se le denomina modular y al causante modulador (en este caso, el objeto), mientras que a la señal emitida por el LED se le denomina portadora, por ser la que va a llevar la información sobre la presencia o no del objeto. A la señal resultante se le llama señal modulada. Además, se puede considerar que la información es digital: por ejemplo, un cero indica la presencia de un objeto y un uno la no presencia. A la propia información (cero presencia del objeto, uno no presencia) se le denomina señal moduladora (figura 21.1).

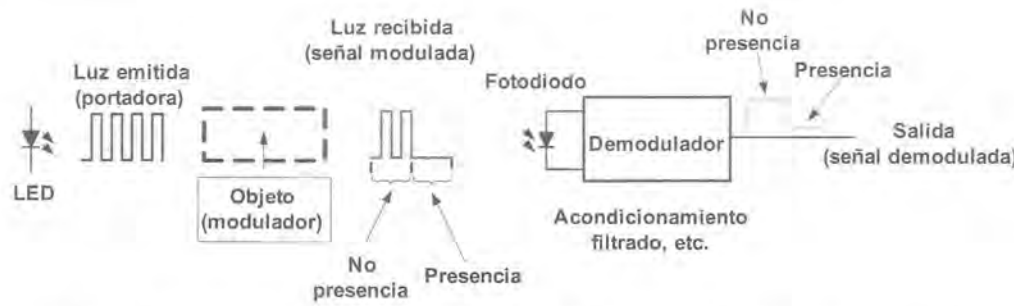


Figura 21.1. Ejemplo de detector de proximidad de barrera en el que se muestran los conceptos básicos relacionados con la modulación y demodulación de señales.

Sin embargo, existen normalmente más fuentes de luz continuas que el propio LED, como puede ser la luz solar y la luz de las lámparas presentes en el recinto en el que se instale la barrera que pueden no verse interrumpidas por la presencia del objeto con lo que podría seguir llegando luz al detector óptico. Resulta necesario distinguir cuál de los haces luminosos corresponde realmente al LED. Para ello, podría manejarse el LED de forma que emitiera una luz de una determinada frecuencia tal que no fuera previsible la presencia de otras fuentes de luz de frecuencia igual o próxima. Por tanto, puede resultar interesante que la portadora tenga, por ejemplo, una frecuencia fácilmente distinguible de otras portadoras (como la luz solar) por procedimientos electrónicos. De todas formas, al fotodiodo llegarían todas las señales (la de la luz solar, las lámparas y el LED). Sería necesario, a partir de todas estas señales, obtener la señal buscada. A esta acción se le denomina demodular la señal.

De todas formas, el presente capítulo está dedicado a la modulación y demodulación de señales dentro del ámbito de los sistemas de comunicaciones. Supóngase, por ejemplo, que se desean enviar por un mismo cable varias señales procedentes de diferentes sensores. Siguiendo el mismo argumento que el anterior, a cada una de ellas podría asignársele una portadora de diferente frecuencia, transmitirlas todas por el mismo cable y luego demodular la señal resultante y recobrar cada una de las señales primitivas. Un ejemplo muy similar es de sobra conocido por todos: la transmisión por radio. Así, en la transmisión por radio en AM o FM, varias señales están compartiendo un mismo medio de transmisión (el espacio) para lo que se modulan y se demodulan de la forma conveniente.

La modulación de señales permite, por tanto, multiplexar en frecuencia varias señales sobre un mismo canal de transmisión y que la transmisión se realice con una mayor inmunidad frente a las interferencias. Además, en los sistemas de comunicaciones por radio, el tamaño de las antenas está directamente relacionado con la frecuencia de la señal que se pretende transmitir por lo que empleando portadores de frecuencias elevadas se pueden emplear antenas más pequeñas.

El esquema básico de un sistema de comunicaciones se muestra en la figura 21.2. La señal portadora tiene una determinada frecuencia, amplitud y fase. La información que se desea transmitir (la señal moduladora), se incluye en la señal portadora como una variación proporcional de alguno de sus tres parámetros característicos. La señal, así modulada, se transmite por el medio de transmisión que corresponda y el demodulador recupera la señal original.

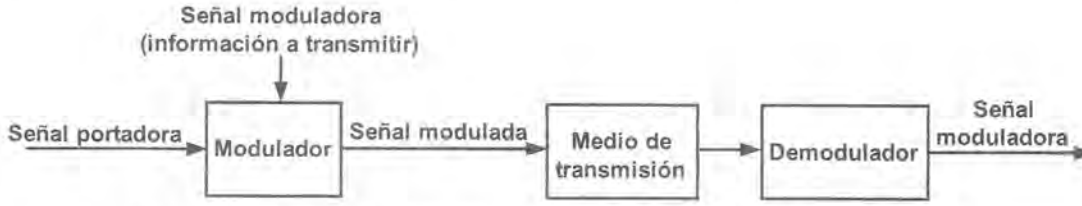


Figura 21.2. Esquema básico de un sistema de modulación y demodulación.

En función de si la portadora y la moduladora son señales analógicas o digitales y de si la información de la moduladora se va a incluir en la amplitud, en la frecuencia o en la fase de la portadora, se distinguen los tipos de modulación:

- Modulación con portadora analógica y moduladora analógica:
  - Modulación de amplitud, AM (*Amplitude Modulation*).
  - Modulación de frecuencia, FM (*Frequency Modulation*).
  - Modulación de fase, PM (*Phase Modulation*):

- Modulación con portadora analógica y moduladora digital.
  - Modulación de amplitud, ASK (*Amplitude Shift Keying*).
  - Modulación de frecuencia, FSK (*Frequency Shift Keying*).
  - Modulación de fase, PSK (*Phase Shift Keying*).
- Modulación con portadora digital y moduladora digital:
  - Modulación de amplitud, PAM (*Pulse Amplitude Modulation*).
  - Modulación de frecuencia, PWM (*Pulse Width Modulation*).
  - Modulación de fase, PPM (*Pulse Position Modulation*).
  - Un caso especial es la modulación por pulsos codificados, PCM (*Pulse Code Modulation*).

Desde el punto de vista de los sistemas de medida, el conocimiento de cada uno de estos tipos de modulación tiene gran interés, por ejemplo, para los sistemas de telemetría por lo que, a continuación, se procederá a un breve estudio de los mismos.

21.1 Modulación con portadora analógica y moduladora analógica

La modulación con portadora analógica y moduladora analógica tiene, hoy en día, escaso interés dentro del ámbito de los sistemas de medición. No obstante, se estudiarán a continuación los diferentes métodos de modulación por este procedimiento (AM, FM y PM) puesto que la teoría en que se sustentan es idéntica a la del resto de los sistemas de modulación que no dejan de ser casos particulares de este método genérico.

21.1.1 MODULACIÓN AM

La modulación AM (*Amplitude Modulation*) corresponde a una modulación con portadora analógica y moduladora analógica. La portadora, V_p, es una señal de una determinada frecuencia y amplitud (figura 21.3a):

$$V_p(t) = A_p \cos(\omega_p t) \tag{21.1}$$

Supóngase, por el momento, que la señal moduladora V_m es una señal de una única frecuencia mucho menor que la de la portadora (figura 21.3b):

$$V_m(t) = A_m \cos(\omega_m t) \tag{21.2}$$

Esta modulación consiste en transmitir la información de la moduladora como una variación de la amplitud de la portadora. Una posibilidad es sumar la amplitud de la portadora y de la moduladora con lo que la señal modulada, V_{pm} es (figura 21.3c):

$$V_{pm} = \left[A_p + A_m \cos(\omega_m t)\right] \cos(\omega_p t) \tag{21.3}$$

Como se puede observar en la figura 21.3, la envolvente de la señal modulada es la propia moduladora por lo que la demodulación puede realizarse, por ejemplo, mediante un circuito electrónico que obtenga la envolvente de la señal modulada. Para ello debe cumplirse, además, que la amplitud de la señal moduladora sea inferior a la de la portadora; de lo contrario, se produciría una sobremodulación (figura 21.4). A la relación entre la amplitud de la moduladora y de la portadora se le denomina índice de modulación, m_a:

$$m_a = \frac{A_m}{A_p} \tag{21.4}$$

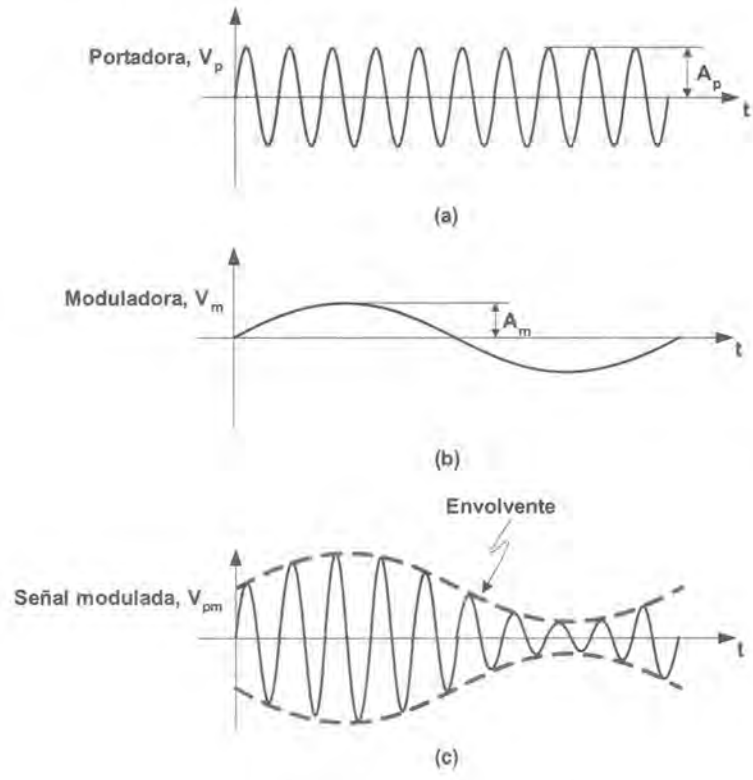


Figura 21.3. Modulación AM. (a) Señal portadora; (b) señal moduladora; (c) señal modulada.

Por tanto, para que no se produzca la sobremodulación, el índice de modulación debería ser menor o igual que uno.

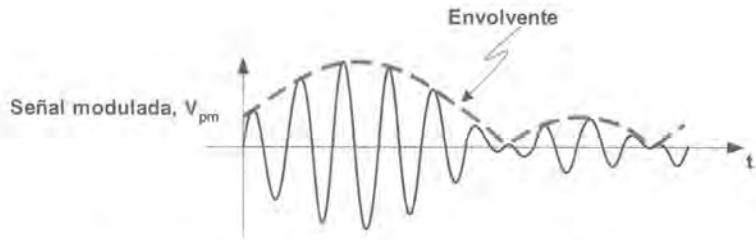


Figura 21.4. Modulación AM, señal sobremodulada.

Espectro

El estudio del espectro de frecuencias de la señal modulada tiene un gran interés puesto que permite conocer cuál es el ancho de banda que se necesita para realizar la transmisión de la señal. En el caso de la modulación AM, resulta muy sencillo determinar el espectro. Así, de la ecuación (21.3) se obtiene:

$$V_{pm}(t) = A_p \cos(\omega_p t) + A_m \cos(\omega_m t)\cos(\omega_p t) \qquad (21.5)$$

o lo que es lo mismo[1]:

$$V_{pm}(t) = A_p \cos(\omega_p t) + \frac{A_m}{2}\cos\left[(\omega_p + \omega_m)t\right] + \frac{A_m}{2}\cos\left[(\omega_p - \omega_m)t\right] \qquad (21.6)$$

Es decir, el espectro está compuesto por tres señales (figura 21.5):

- La propia portadora.
- Una señal de la mitad amplitud que la moduladora y cuya frecuencia es la de la portadora, f_p, menos la de la moduladora, f_m.
- Otra señal idéntica a la anterior salvo que la frecuencia es la de la portadora más la de la moduladora.

Estas dos últimas señales corresponden a lo que denominan bandas laterales puesto que gráficamente están situadas a "ambos lados" de la señal portadora.

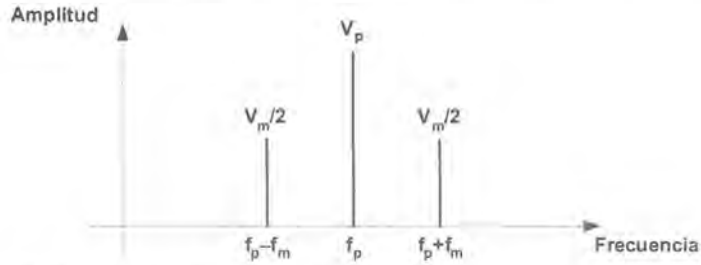


Figura 21.5. Espectro de una señal modulada en AM cuando la moduladora es una señal de una única frecuencia.

Sin embargo, la señal moduladora no tiene en general una única frecuencia. Por ejemplo, en aplicaciones de audio, la señal moduladora podría tener un espectro en frecuencias comprendido entre 20 Hz y 20 kHz. Cada una de estas frecuencias daría lugar a la aparición de las correspondientes frecuencias en las bandas laterales. Denominando $f_{m\text{-}mín}$ y $f_{m\text{-}máx}$ a las frecuencias mínima y máxima de la moduladora, respectivamente, el espectro completo de la señal modulada estaría comprendido entre $f_p \pm f_{m\text{-}máx}$ (figura 21.6), es decir, el ancho de banda de la señal modulada sería de dos veces la frecuencia máxima de la señal moduladora.

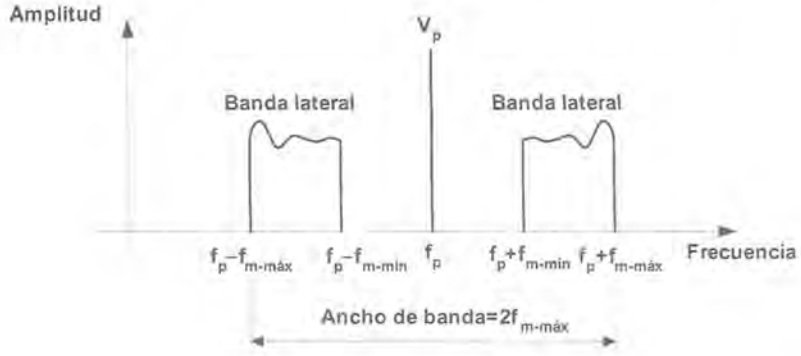


Figura 21.6. Espectro de una señal modulada en AM cuando la moduladora es una señal cuyas frecuencias están comprendidas entre $f_{m\text{-}mín}$ y $f_{m\text{-}máx}$.

[1] $\cos(a)\cos(b) = \frac{1}{2}\cos(a+b) + \frac{1}{2}\cos(a-b)$

Por ejemplo, la radiodifusión comercial en AM se realiza entre 531 kHz y 1602 kHz. Para que puedan coexistir un elevado número de emisoras, se ha acordado limitar el ancho de banda para cada una de ellas a 9 kHz con lo que la señal de audio debe ser de una frecuencia de menos de 4,5 kHz. Evidentemente, con este ancho de banda para la señal de audio, la calidad del sonido es relativamente pobre.

EJERCICIO 21.1

Se desean transmitir dos señales de frecuencias comprendidas entre 10 Hz y 10 kHz moduladas en AM convencional. Para una de las señales se empleará una portadora de 1000 kHz, ¿se podría utilizar una portadora de 1005 kHz para la segunda señal?

SOLUCIÓN:

No. El espectro ocupado por la primera señal modulada estará comprendido entre 990 y 1010 kHz por lo que, si la segunda señal se transmite alrededor de 1005 kHz, se produciría la mezcla de ambas. Por este motivo, para la segunda señal será necesario emplear una portadora de, al menos, 1020 kHz o inferior a 980 kHz.

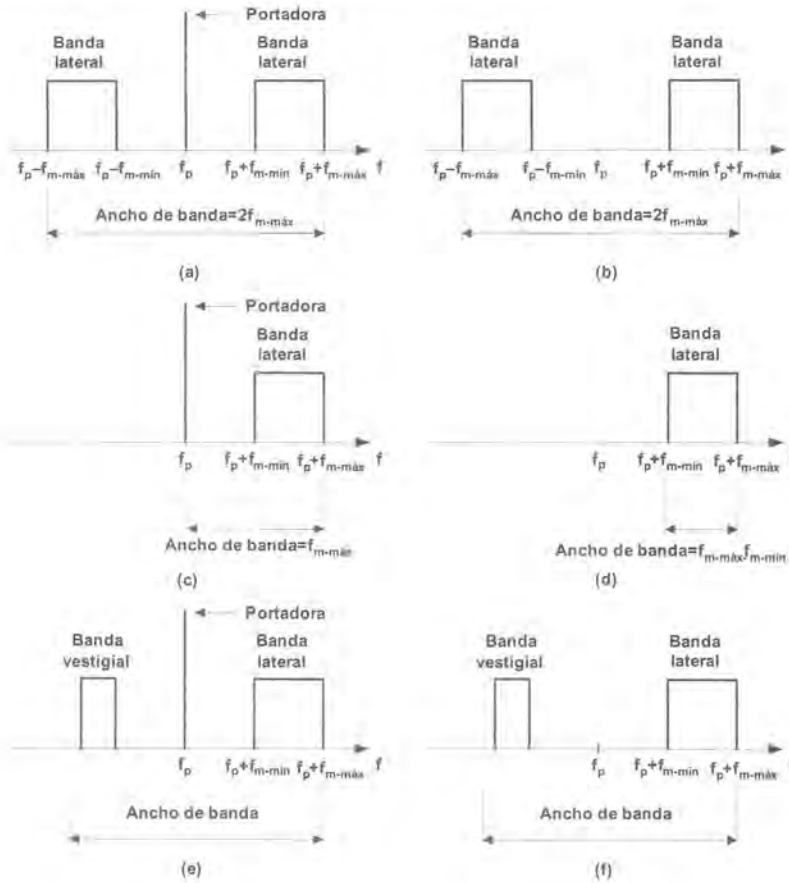


Figura 21.7. Representación espectral de diferentes tipos de modulación AM. (a) Convencional; (b) doble banda lateral con portadora suprimida; (c) banda lateral única con portadora; (d) banda lateral única con portadora suprimida; (e) banda vestigial con portadora; (f) banda vestigial con portadora suprimida.

En el espectro de frecuencias de la señal modulada (figura 21.6) se puede observar que está contenida la propia señal portadora, que no contiene información alguna y por lo tanto, se podría eliminar de la señal modulada. Esto tiene gran importancia, porque la potencia necesaria para transmitir la señal disminuiría considerablemente. Pero además, la información de las bandas laterales es redundante por lo que se podría eliminar una de ellas sin ningún inconveniente aunque la demodulación de la señal podría ser más complicada de llevar a cabo. La modulación AM puede realizarse, por lo tanto, de diferentes modos distinguiéndose, entre otros, los siguientes (figura 21.7):

- AM en doble banda lateral (*AM-DSB, Double Side Band*): se transmite la señal modulada completa, es decir, las dos bandas laterales y la portadora.
- AM con portadora suprimida (*AM-DSB-SC, Double Side Band-Suppressed Carrier*): presenta la ventaja de disminuir notablemente la potencia necesaria para la transmisión.
- AM en banda lateral única con portadora (*AM-SSB-TC, Single Side Band-Transmitted Carrier*): la ventaja reside en que el ancho de banda ocupado es menor.
- AM en banda lateral única (*AM-SSB-SC, Single Side Band-Suppressed Carrier*): combina las ventajas de las dos anteriores.
- AM en banda lateral vestigial (*AM-VSB, Vestigial Side Band*): se transmite la portadora, una banda lateral y una porción de la otra banda. Se utiliza en televisión para la transmisión de señales de vídeo. En ocasiones, también se suprime la portadora (VSB-SC). Presenta la ventaja de que el ancho de banda es reducido, aunque no tanto como en banda lateral, pero la demodulación es mucho más sencilla de realizar.

21.1.2 MODULACIÓN FM

Mientras que el objetivo de la modulación AM era transmitir la información de la moduladora como una variación de la amplitud de la portadora, en la modulación FM el objetivo es transmitir la información como una variación de la frecuencia instantánea.

En primer lugar resulta necesario aclarar qué significa el término frecuencia instantánea. Por ejemplo, una señal cosenoidal se representa en términos generales por la función:

$$F(\theta) = A \cdot \cos(\theta) \tag{21.7}$$

donde θ es la fase. Cuando la fase varía de forma lineal con el tiempo se tiene que:

$$\theta = \omega t \tag{21.8}$$

denominándose velocidad angular o pulsación, ω, a la constante de proporcionalidad (representa en el fondo la rapidez con la que varía el ángulo):

$$\omega = \frac{d\theta}{dt} \tag{21.9}$$

y la frecuencia es la velocidad angular dividido entre 2π. En la modulación FM se tratará de modificar la frecuencia instantánea de forma proporcional al valor instantáneo de la señal moduladora. Si denominamos k_f a la razón de proporcionalidad y suponiendo, como al estudiar la modulación AM, que la señal moduladora es de una única frecuencia, la frecuencia instantánea de la señal modulada es:

$$f_{pm}(t) = f_p + k_f A_m \cos(\omega_m t) \tag{21.10}$$

Por lo tanto, la fase de la señal es:

$$\theta(t) = \int \omega dt = 2\pi \int f dt = 2\pi \iint \left[f_p + k_f A_m \cos(\omega_m t) \right] dt \tag{21.11}$$

Operando, se obtiene:

$$\theta(t) = 2\pi f_p t + \frac{2\pi k_f A_m}{\omega_m} \operatorname{sen}(\omega_m t) = \omega_p t + \frac{2\pi k_f A_m}{\omega_m} \operatorname{sen}(\omega_m t) \tag{21.12}$$

Así pues, la expresión de la señal modulada será:

$$V_{pm}(t) = A_p \cos\left[\omega_p t + \frac{2\pi k_f A_m}{\omega_m} \operatorname{sen}(\omega_m t) \right] \tag{21.13}$$

Por otra parte, observando la ecuación (21.10), se puede concluir que la frecuencia instantánea de la señal modulada está comprendida entre $f_p \pm k_f A_m$. Se denomina desviación de frecuencia Δf al valor

$$\Delta f = k_f A_m \tag{21.14}$$

y, en la modulación FM, se denomina índice de modulación m_f, a la relación entre la desviación de la frecuencia y la frecuencia de la moduladora:

$$m_f = \frac{\Delta f}{f_m} \tag{21.15}$$

por lo que la ecuación (21.13), se puede expresar como:

$$V_{pm} = A_p \cos\left[\omega_p t + m_f \operatorname{sen}(\omega_m t) \right] \tag{21.16}$$

En la figura 21.8, se puede observar el aspecto de la señal modulada.

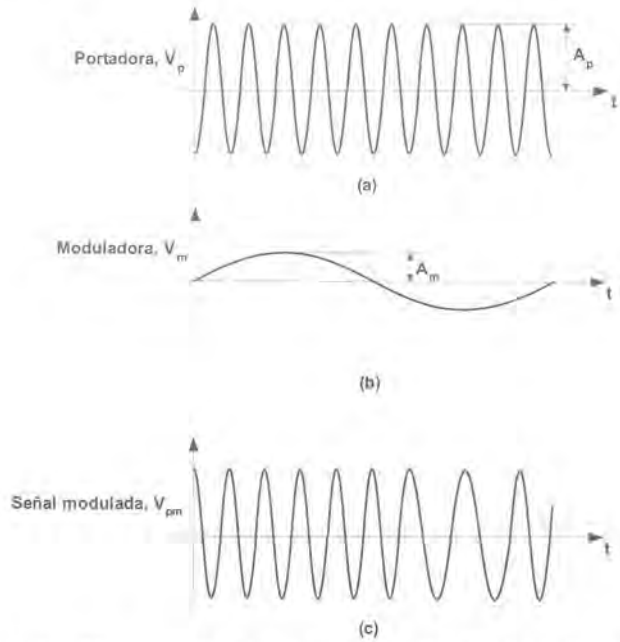


Figura 21.8. Modulación FM. (a) Señal portadora; (b) señal moduladora; (c) señal modulada.

Espectro

La obtención de las frecuencias que componen la señal modulada puede realizarse a partir del desarrollo en serie de Fourier. El análisis es más complejo que en el caso de la modulación AM. Una vez realizado se obtiene que la señal modulada es:

$$V_{pm} = A_p \sum_{n=-\infty}^{\infty} \left[J_n(m_f) \cos(\omega_p + n\omega_m) \right] \tag{21.17}$$

donde $J_n(m_f)$ son los coeficientes de Bessel (tabla 21.1):

$$J_n(m_f) = \left(\frac{m_f}{2}\right)^n \left\{ \sum_{k=0}^{\infty} \left[\frac{(m_f/2)^{2k}}{k!(n+k)!}(-1)^k \right] \right\} \tag{21.18}$$

A la vista de la ecuación (21.17) se puede observar que el espectro en frecuencias está compuesto por infinitas señales cuyas frecuencias son $f_p \pm nf_m$ donde n=0,1,2..., y cuyas amplitudes son $A_pJ_n(m_f)$. En la figura 21.9, se puede observar el espectro para diferentes índices de modulación.

m_f	J_0	J_1	J_2	J_3	J_4	J_5	J_6	J_7	J_8	J_9	J_{10}	J_{11}	J_{12}	J_{13}	J_{14}	J_{15}
0	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
0,5	0,938	0,242	0,031	-	-	-	-	-	-	-	-	-	-	-	-	-
1	0,765	0,44	0,115	0,02	-	-	-	-	-	-	-	-	-	-	-	-
1,5	0,512	0,558	0,232	0,061	0,012	-	-	-	-	-	-	-	-	-	-	-
2	0,224	0,577	0,353	0,129	0,034	-	-	-	-	-	-	-	-	-	-	-
2,5	-0,048	0,497	0,446	0,217	0,074	0,02	-	-	-	-	-	-	-	-	-	-
3	-0,26	0,339	0,486	0,309	0,132	0,043	0,011	-	-	-	-	-	-	-	-	-
3,5	-0,38	0,137	0,459	0,387	0,204	0,08	0,025	-	-	-	-	-	-	-	-	-
4	-0,397	-0,066	0,364	0,43	0,281	0,132	0,049	0,015	-	-	-	-	-	-	-	-
4,5	-0,321	-0,231	0,218	0,425	0,348	0,195	0,084	0,03	-	-	-	-	-	-	-	-
5	-0,178	-0,328	0,047	0,365	0,391	0,261	0,131	0,053	0,018	-	-	-	-	-	-	-
6	0,151	-0,277	-0,243	0,115	0,358	0,362	0,246	0,13	0,057	0,021	-	-	-	-	-	-
7	0,3	0	-0,301	-0,168	0,158	0,348	0,339	0,234	0,128	0,059	0,024	-	-	-	-	-
8	0,172	0,235	-0,113	-0,291	-0,105	0,186	0,338	0,321	0,223	0,126	0,061	0,026	-	-	-	-
9	-0,09	0,245	0,145	-0,181	-0,265	-0,055	0,204	0,327	0,305	0,215	0,125	0,062	0,027	0,011	-	-
10	-0,246	0,043	0,255	0,058	-0,22	-0,234	-0,014	0,217	0,318	0,292	0,207	0,123	0,063	0,029	0,012	-

Tabla 21.1. Coeficientes de Bessel para diferentes índices de modulación (se incluyen únicamente los que tienen valores mayores que 0,01).

El ancho de banda ocupado en FM es en principio infinito; no obstante, como se puede observar en la figura 21.9, las principales componentes de la señal modulada aparecen alrededor de la frecuencia de la portadora. Las componentes que están muy alejadas se pueden despreciar puesto que su valor es muy pequeño (en general, se desprecian las componentes de menos del 1% de la amplitud de la portadora). Así, el ancho de *banda ocupado es:*

$$BW = 2Nf_m \tag{21.19}$$

donde N es el número del coeficiente de Bessel de mayor orden que sea mayor que 0,01. Cuanto mayor sea el índice de modulación mayor será N y, por tanto, mayor será el ancho de banda ocupado. Puede observarse en la figura 21.9 que, cuando el índice de modulación es pequeño, el ancho de banda es pequeño (similar al de la modulación AM) por lo que se denomina modulación FM de banda estrecha.

Por otra parte, la regla de Carson ofrece un método sencillo para calcular, aproximadamente, el ancho de banda en función del índice de modulación:

$$BW \approx 2f_m(1+m_f) \tag{21.20}$$

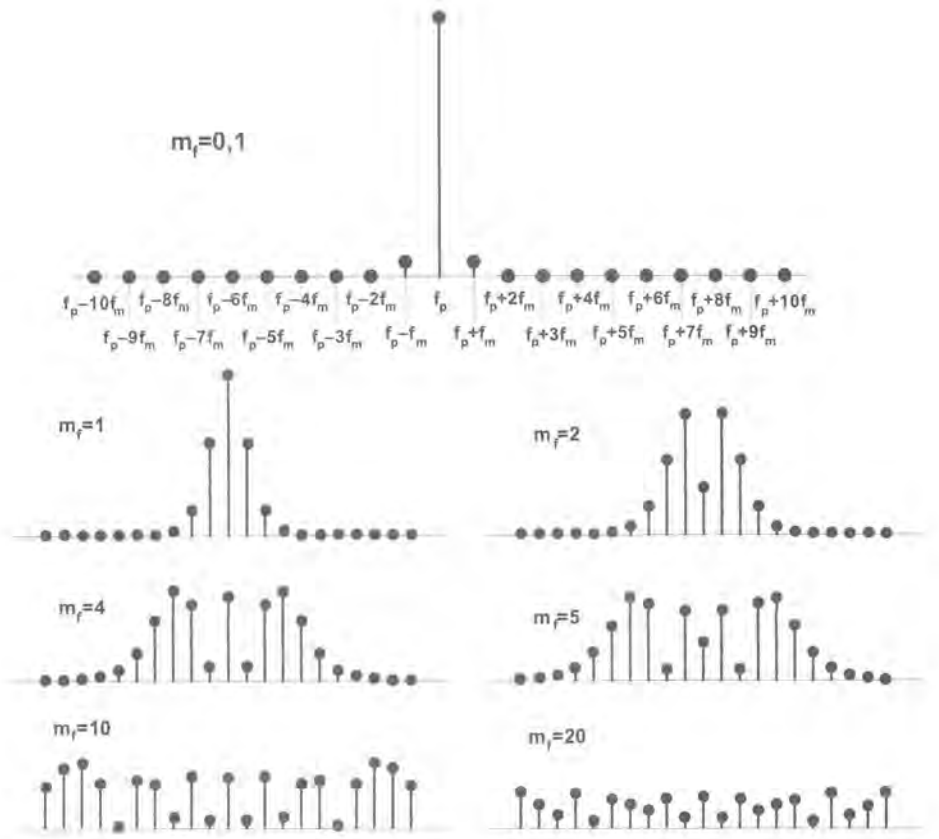


Figura 21.9. Espectro en frecuencias de una señal modulada en FM para diferentes índices de modulación cuando la moduladora es una señal senoidal de una única frecuencia.

Esta aproximación ofrece buenos resultados cuando el índice de modulación es mayor de 20.

En la radiodifusión FM comercial se ha asignado la banda comprendida entre 88 MHz y 108 MHz. Cada emisora tiene permitida una desviación de frecuencia máxima de 75 kHz y puede ocupar un ancho de banda de 200 kHz. La señal de audio que se puede transmitir es de una frecuencia máxima de 15 kHz. Por tanto, el índice de modulación es igual a 5. Con este índice de modulación existen 8 componentes significativas por lo que el ancho de banda ocupado es de 240 kHz, mayor que el permitido. No obstante, son pocas las frecuencias de 15 kHz y, además, las emisoras se separan más de los 200 kHz asignados, por seguridad.

La ventaja fundamental de la modulación FM frente a la modulación AM es que resulta mucho menos sensible a las interferencias. En general, las interferencias modifican la amplitud de la señal pero es mucho más difícil que la frecuencia se vea afectada. Sin embargo, presenta como inconvenientes que el ancho de banda ocupado es mayor y que la modulación y demodulación de la señal es más complicada, lo que encarece los sistemas.

21.1.3 MODULACIÓN PM

En la modulación PM se varía la fase de la señal portadora de forma proporcional a la señal moduladora. Denominando k_p a la razón de proporcionalidad:

$$\theta(t) = \omega_p t + k_p A_m \cos(\omega_m t) \tag{21.21}$$

Por lo tanto, la señal modulada, es (figura 21.10):

$$V_{pm}(t) = A_p \cos\left[\omega_p t + k_p A_m \cos(\omega_m t)\right] \tag{21.22}$$

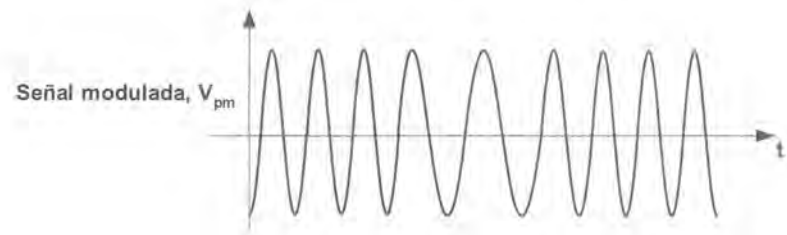


Figura 21.10. Aspecto de una señal modulada en PM (la portadora y la moduladora son las mismas que las mostradas en la figura 21.8).

Como se puede observar, la expresión es muy similar a la que corresponde a la modulación FM (véase la ecuación (21.16) en la que se puede observar que la modulación FM se convierte en una modulación de la fase de la portadora). La modulación FM y PM están estrechamente relacionadas y conducen a resultados muy similares. Las dos se denominan, también, modulaciones angulares. En la modulación PM el índice de modulación, m_p, se define como:

$$m_p = k_p A_m \tag{21.23}$$

y el espectro se obtiene, al igual que en la modulación FM, a partir de los coeficientes de Bessel.

21.2 Modulación con portadora analógica y moduladora digital

21.2.1 MODULACIÓN ASK

En la modulación ASK, la portadora es una señal analógica (figura 21.11a), mientras que la moduladora es digital por lo que únicamente puede tener dos valores ("cero" o "uno") (figura 21.11b). No deja de ser un caso particular de la modulación AM. La señal modulada se obtiene directamente sin más que asignar dos amplitudes diferentes para la señal modulada en función de los dos valores de la moduladora (figura 21.11c).

$$\begin{aligned} V_{pm}(t) &= k_1 V_p(t) \text{ si } V_m = 1 \\ V_{pm}(t) &= k_2 V_p(t) \text{ si } V_m = 0 \end{aligned} \tag{21.24}$$

donde k_1 y k_2 son constantes. En lo sucesivo se considerará que $k_1 = 1$. También es muy habitual asignar a k_2 el valor 0, lo que se conoce como modulación OOK (*On-Off Keying*). En cualquier caso, como se puede observar en la figura 21.11c, la envolvente de la señal modulada refleja la señal moduladora al igual que en AM.

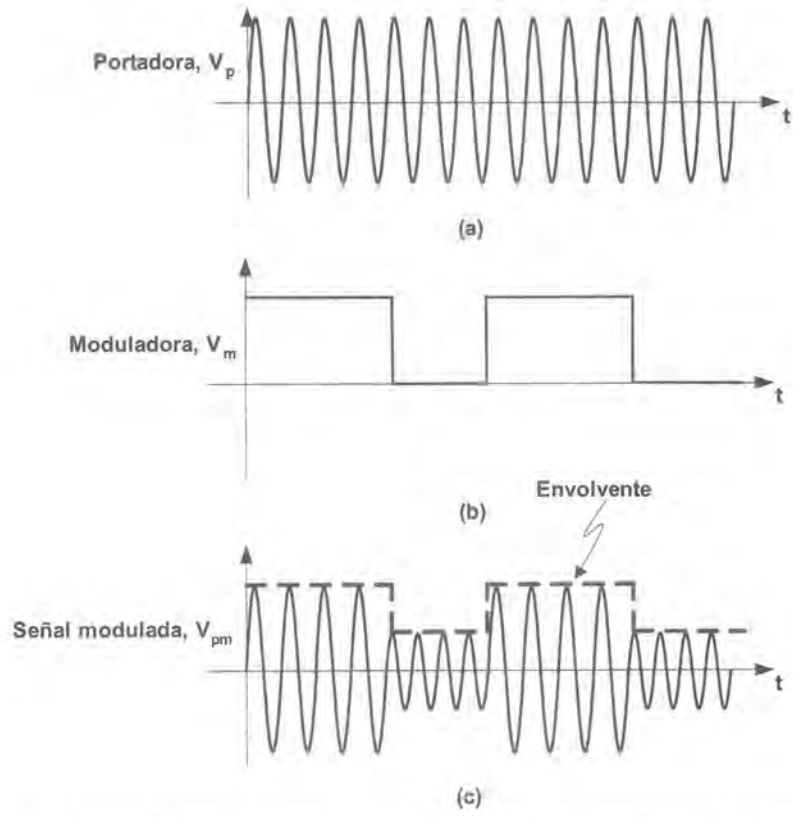


Figura 21.11. Modulación ASK. (a) Señal portadora; (b) señal moduladora; (c) señal modulada.

Espectro

Puesto que la modulación ASK es idéntica a la AM, salvo que la señal moduladora sólo puede tomar dos valores, el espectro en frecuencias es similar. Es decir, el ancho de banda ocupado es dos veces el ancho de banda de la señal moduladora. Sin embargo, ¿cuál es el ancho de banda de la señal moduladora?

Supóngase que se desean transmitir B bits por segundo: la máxima frecuencia de esta señal correspondería al caso en que los bits a transmitir fueran alternativamente "unos" y "ceros" (figura 21.12a). La señal moduladora tendría, por tanto, un periodo igual a 2/B y su frecuencia fundamental seria $f_m = ½B$. Sin embargo, el desarrollo en serie de Fourier de una señal cuadrada de estas características es:

$$V_m(t) = \frac{A_m}{2} + \sum_{n=1}^{\infty}\left(\frac{A_m}{2}\frac{\operatorname{sen}\left(\frac{n\pi}{2}\right)}{\frac{n\pi}{2}}\right)\cos\left(n2\pi f_m t\right) \qquad (21.25)$$

e incluye infinitas frecuencias cuyos valores son nf_m donde n = 0,1,2,... (figura 21.12b). Por lo tanto, el ancho de banda de la señal modulada es infinito (figura 21.12c). No obstante, se pueden despreciar las componentes poco significativas; de hecho es suficiente con la componente fundamental. Así, el ancho de banda ocupado abarca desde $\pm f_m$ por lo que es igual a la velocidad de transmisión, B. Si por ejemplo, se deseara transmitir una señal de 1200 bps (bits por segundo), el ancho de banda ocupado sería de 1200 Hz.

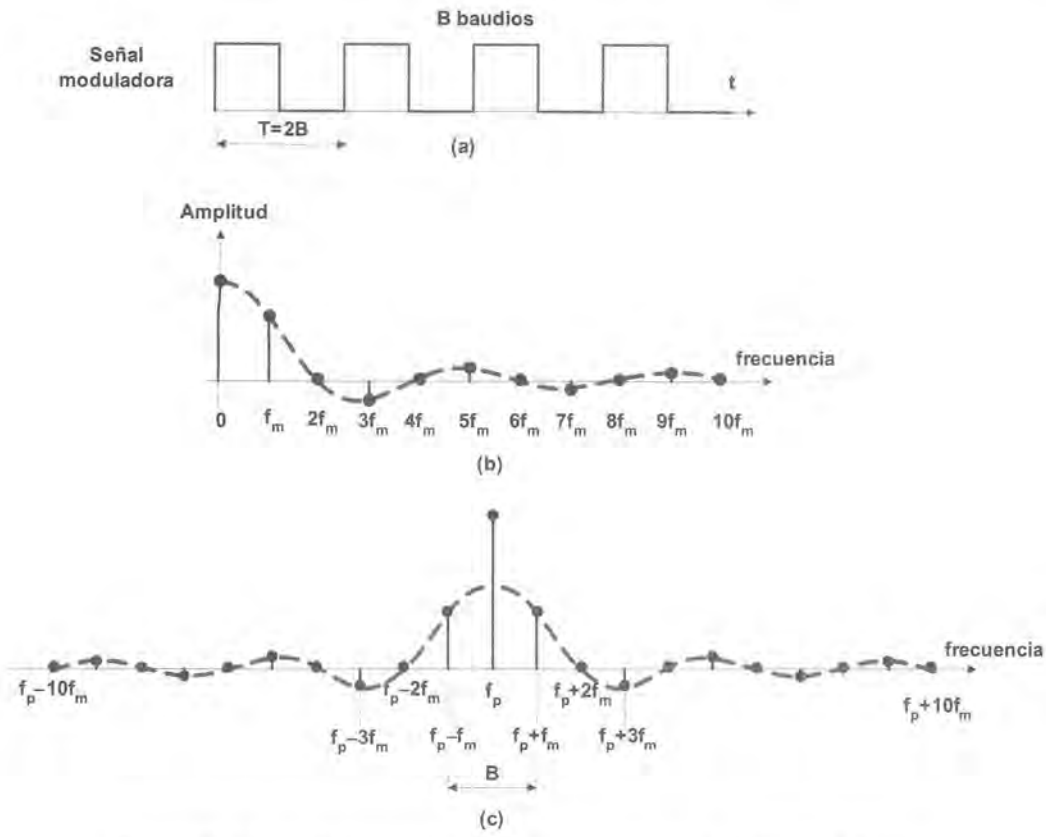


Figura 21.12. (a) Señal moduladora; (b) espectro de la señal moduladora; (c) espectro de la señal modulada.

Circuitos para realizar la modulación y la demodulación

La modulación ASK es muy sencilla de realizar. Una posibilidad consiste simplemente en generar una portadora con un oscilador y seleccionar, con un multiplexador analógico gobernado por la señal moduladora, entre la señal portadora con o sin amplificar (figura 21.13).

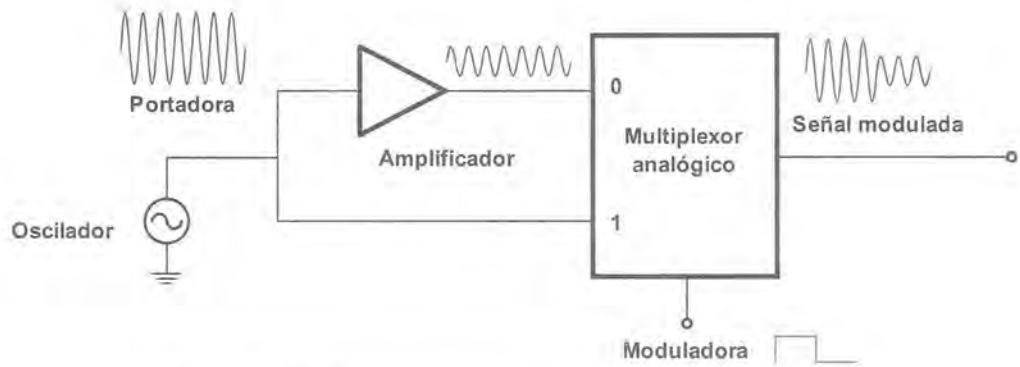


Figura 21.13. Diagrama de bloques de un modulador de ASK.

La demodulación puede realizarse sin más que obtener la envolvente de la señal. Para ello se puede utilizar un circuito como el de la figura 21.14 y comparar posteriormente la envolvente con una señal de referencia.

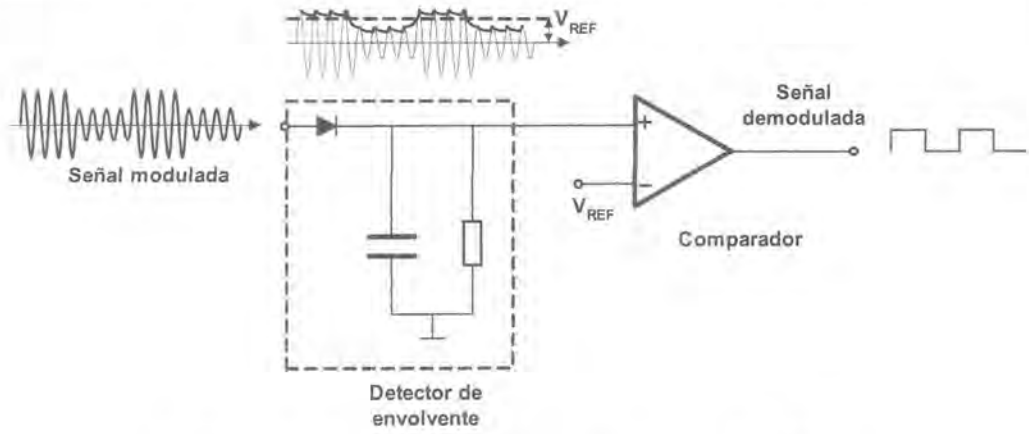


Figura 21.14. Esquema básico de un demodulador ASK.

21.2.2 MODULACIÓN FSK

La modulación FSK consiste en asignar dos frecuencias diferentes para los dos valores de la señal modulada (figura 21.15). Por ejemplo, si la portadora V_p tiene una pulsación ω_p, la señal modulada, V_{pm} se puede expresar como:

$$\begin{aligned} V_{pm}(t) &= A_p \operatorname{sen}(\omega_p t) \quad \text{si} \quad V_m = 0 \\ V_{pm}(t) &= A_p \operatorname{sen}(k\omega_p t) \quad \text{si} \quad V_m = 1 \end{aligned} \tag{21.26}$$

donde k es una constante. O lo que es lo mismo, se asigna una frecuencia al "cero" f_0 (igual a la frecuencia de la portadora, f_p, en este caso) y una frecuencia al uno, f_1 (igual a kf_p).

Espectro

La obtención del espectro se realiza como en el caso de la modulación FM. Sin embargo, podemos aprovechar los resultados ya obtenidos.

Si deseamos transmitir una señal de B bps, la frecuencia fundamental, f_m, de la moduladora es como máximo ½B. Si tenemos en cuenta las ecuaciones (21.10) y (21.14), la frecuencia de la señal modulada podrá variar como máximo entre $\pm\Delta f$ (Δf es la desviación de frecuencia). Si hemos asignado al cero, la frecuencia f_0 y al uno la frecuencia f_1, la desviación de frecuencia será:

$$\Delta f = \frac{|f_1 - f_0|}{2} \tag{21.27}$$

es decir, la modulación FSK, equivale a tener una portadora cuya frecuencia, f_p, es la frecuencia central de las asignadas de forma que la frecuencia instantánea de la señal modulada puede variar entre $f_p \pm \Delta f$. Por otra parte, el índice de modulación es:

$$m_f = \frac{\Delta f}{f_m} \tag{21.28}$$

por lo que se puede obtener el espectro sin más que determinar los coeficientes de Bessel.

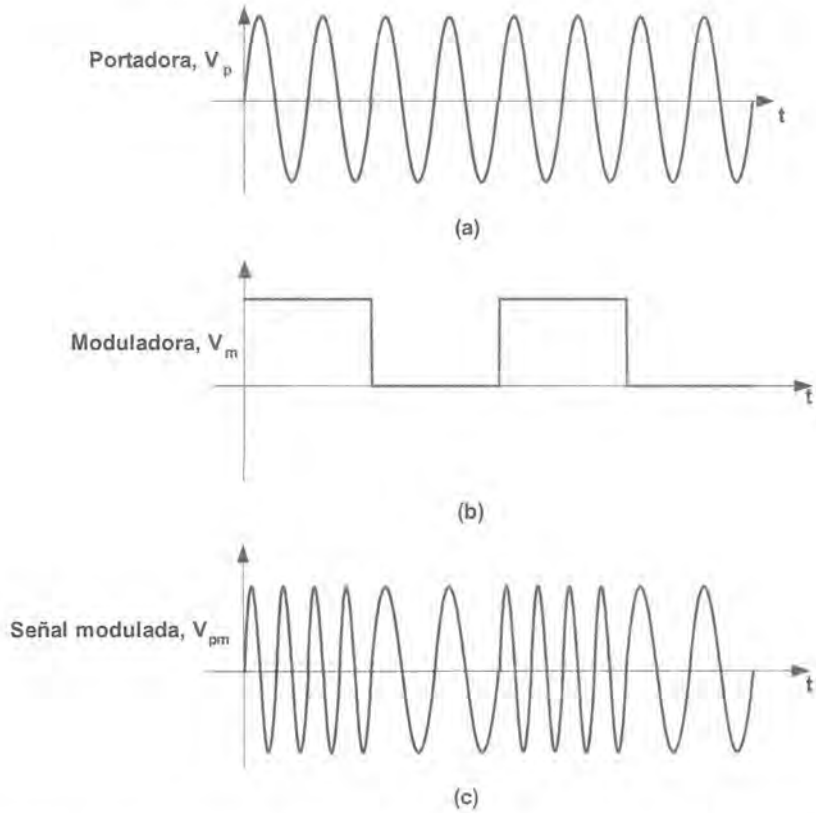


Figura 21.15. Modulación FSK. (a) Señal portadora; (b) señal moduladora; (c) señal modulada.

EJERCICIO 21.2

Se desea realizar un modulador FSK de forma que se le asigne al cero la frecuencia de 1300 Hz y al uno 2100 Hz. Suponiendo que la tasa de bits de entrada es de 1200 bps, determine el ancho de banda necesario.

SOLUCIÓN:

La desviación de frecuencia es:

$$\Delta f = \frac{2100 - 1300}{2} = 400 \text{ Hz}$$

y el indice de modulación, teniendo en cuenta que la frecuencia fundamental de la entrada es la mitad de la velocidad de transmisión en bits/s:

$$m_f = \frac{\Delta f}{600} = \frac{400}{600} = 0,667$$

Para este índice de modulación los coeficientes de Bessel son (ecuación 21.18): $J_0(0{,}667) = 0{,}892$; $J_1(0{,}667) = 0{,}315$; $J_2(0{,}667) = 0{,}054$ y los coeficientes de mayor orden son menores de 0,01 por lo que se desprecian. Por tanto, el ancho de banda ocupado, según la ecuación (21.19) es (N = 3):

$$BW = 2 \cdot 2 \cdot 600 = 2400 \text{ Hz}$$

La frecuencia de la portadora, o frecuencia central, es:

$$f_p = 1300 + 400 = 2100 - 400 = 1700 \text{ Hz}$$

por lo que el espectro de la señal modulada está comprendido entre 1700 – 1200 = 500 Hz y 1700 + 1200 = 2900 Hz.

Para realizar la modulación FSK se pueden emplear, por ejemplo, dos osciladores de dos frecuencias diferentes (figura 21.16a), aunque lo más habitual es emplear un oscilador controlado por tensión (VCO) (figura 21.16b).

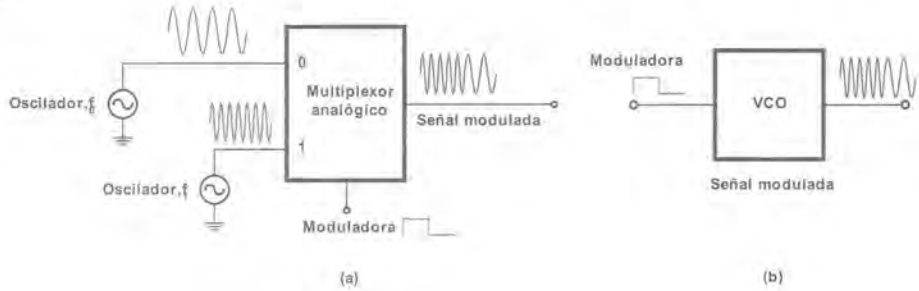


Figura 21.16. Moduladores de FSK.

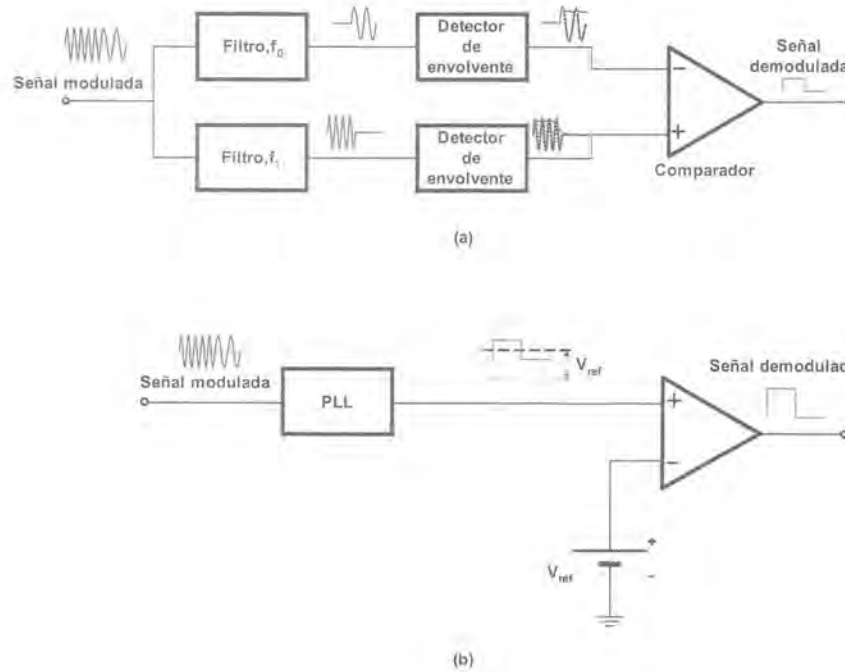


Figura 21.17. Demoduladores de FSK.

La demodulación puede realizarse de múltiples formas. Una posibilidad es emplear dos filtros sintonizados a cada una de las frecuencias a las que se realiza la modulación. Es el esquema denominado "de filtros duales" (figura 21.17a). Otra posibilidad que ofrece mejores resultados consiste en utilizar un PLL (figura 21.17b).

21.2.3 MODULACIÓN PSK

En la modulación de fase, se varía la fase de la señal portadora según el valor de la moduladora. Por ejemplo, se puede asignar fase cero cuando la moduladora es un uno, y fase π cuando vale cero (figura 21.18).

$$V_{pm}(t) = A_p sen(\omega_p t) \;\; si \;\; V_m = 1$$

$$V_{pm}(t) = A_p sen(\omega_p t - \pi) \;\; si \;\; V_m = 0 \qquad (21.29)$$

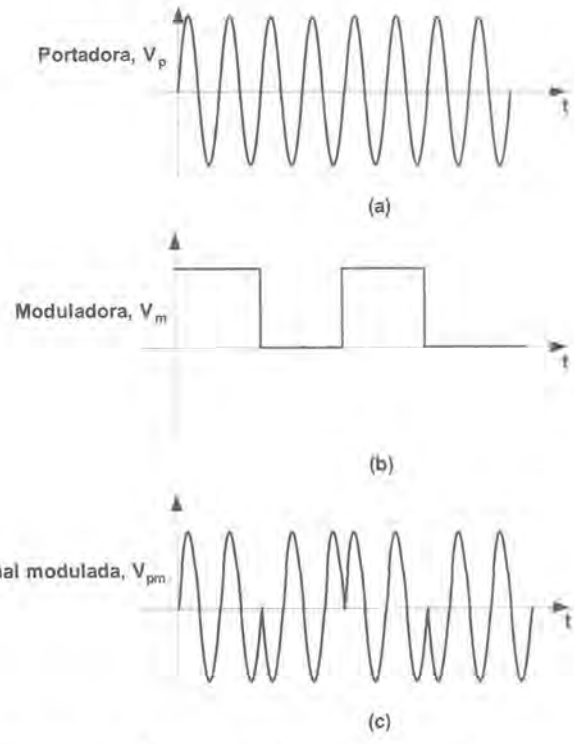


Figura 21.18. Modulación PSK. (a) Señal portadora; (b) señal moduladora; (c) señal modulada.

Si se desea transmitir una señal de B bps, el máximo ancho de banda ocupado tiene lugar cuando la frecuencia de la moduladora es $f_m = ½B$. La señal modulada, en este caso, se obtiene multiplicando la portadora por una señal que toma alternativamente los valores +1 y –1 (véase figura 21.18c). Si se tiene en cuenta únicamente la frecuencia fundamental de la moduladora, la señal modulada es:

$$V_{pm}(t) = V_p sen(\omega_p t) V_m sen(\omega_m t) \qquad (21.30)$$

o lo que es lo mismo:

$$V_{pm}(t) = \frac{1}{2} V_p V_m \cos(\omega_p t - \omega_m t) - \frac{1}{2} V_p V_m \cos(\omega_p t - \omega_m t) \qquad (21.31)$$

es decir, en la señal modulada aparecen dos frecuencias: $f_p + f_m$ y $f_p - f_m$ (no aparece ninguna señal a la frecuencia de la portadora), por lo que el ancho de banda es $2f_m = B$ (figura 21.19).

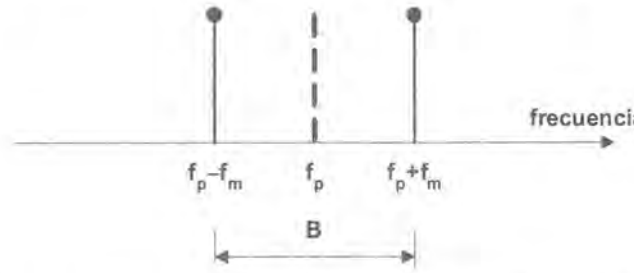


Figura 21.19. Máximo espectro ocupado por la componente fundamental de una señal de B bits por segundo (obsérvese que la portadora aparece suprimida).

EJERCICIO 21.3

Se desea realizar un modulador PSK, empleando una portadora de 1700 Hz. Suponiendo que la tasa de bits de entrada es de 1200 bps, determinar el ancho de banda necesario.

SOLUCIÓN:

El ancho de banda ocupado es igual a 1200 Hz. El espectro de la señal modulada está comprendido (en su mayor parte), entre 1700 – 600 = 1100 Hz y 1700 + 600 = 2300 Hz.

La modulación PSK es muy sencilla de realizar. Un posible esquema, basado en la utilización de un multiplexador, se muestra en la figura 21.20.

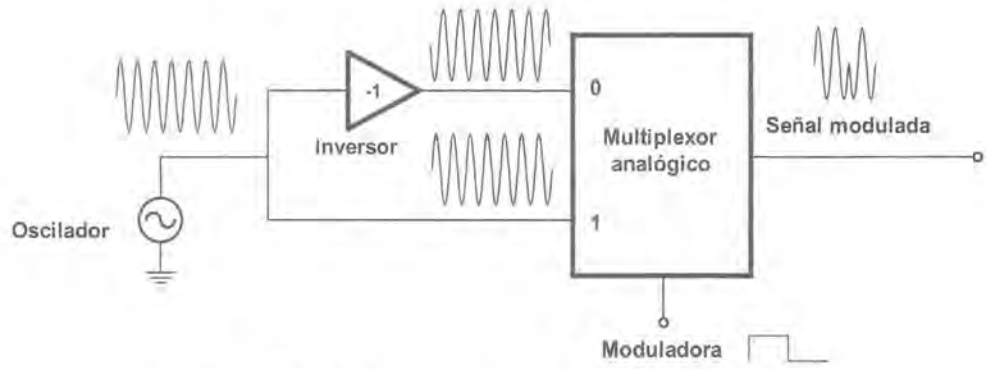


Figura 21.20. Esquema básico de un modulador PSK.

Sin embargo, la demodulación es más complicada. Una posibilidad para realizar la detección de la fase consiste en recuperar la portadora a partir de la señal modulada (la portadora desaparece de la señal modulada). Si posteriormente se multiplican ambas señales, se obtendrá una señal de doble frecuencia de la que se puede obtener fácilmente la información de la fase (figura 21.21).

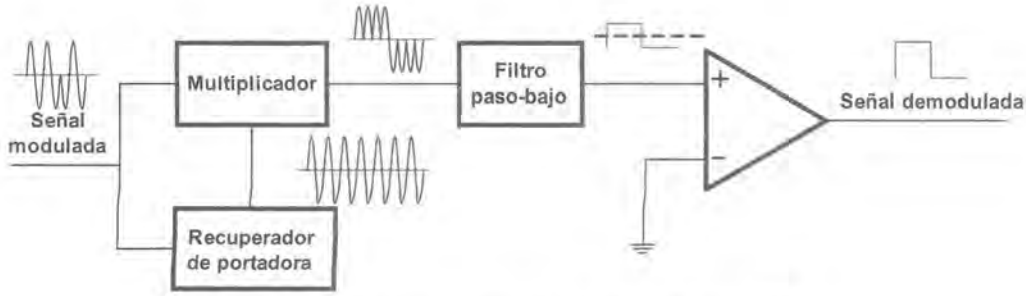


Figura 21.21. Esquema básico de un demodulador PSK.

Modulación QPSK

La modulación por desplazamiento de fase cuaternaria o en cuadratura consiste en asignar a cada conjunto de dos bits una fase diferente. Como existen cuatro posibles combinaciones de dos bits, se necesita asignar cuatro fases diferentes. Por ejemplo (figura 21.22):

- −135° para el 00
- −45° para el 01
- +135° para el 10
- +45° para el 11

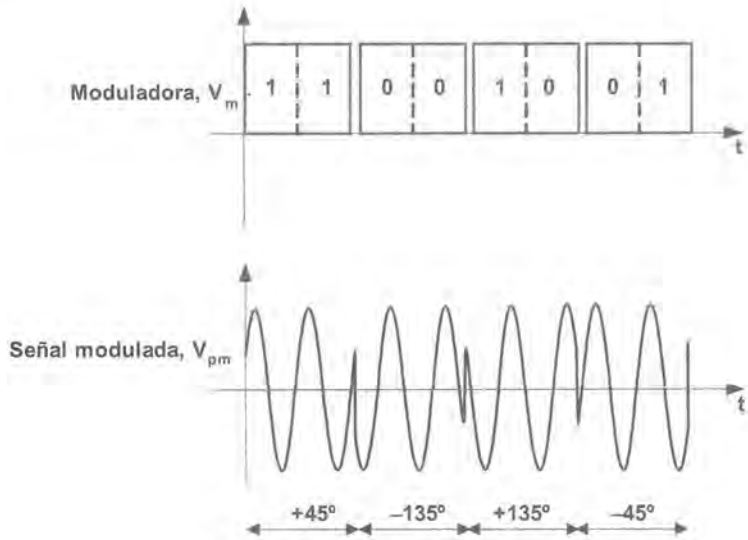


Figura 21.22. Modulación QPSK.

La codificación se suele representar gráficamente mediante un diagrama de constelación (figura 21.23).

Se puede demostrar fácilmente que si se pretenden transmitir B bits por segundo, el ancho de banda ocupado es B/2, es decir, la mitad que en la modulación PSK binaria (BPSK). También cabe la posibilidad de agrupar un mayor número de bits, por ejemplo 3, dando lugar a la modulación PSK de 8 fases y disminuyendo de nuevo el ancho de banda ocupado.

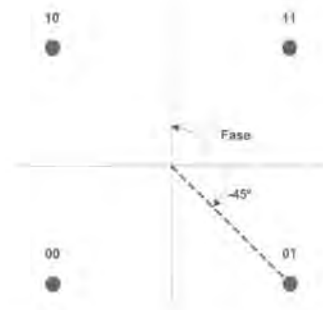


Figura 21.23. Diagrama de constelación de la modulación QPSK.

En ocasiones se combina la modulación PSK con otras. Una combinación muy utilizada es la denominada de amplitud en cuadratura (QAM). Consiste en agrupar 3 bits y asignar a cada grupo una fase-amplitud diferente. Un posible diagrama de constelación sería el que se muestra en la figura 21.24.

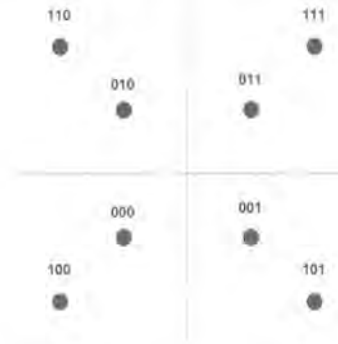


Figura 21.24. Diagrama de constelación de la modulación QAM.

21.3 Modulaciones con portadora digital y moduladora analógica

En este tipo de modulaciones, la señal digital portadora, cuya frecuencia es f_p (se denomina frecuencia de muestreo), ve modificada, como en el caso de las modulaciones tratadas hasta el momento, la amplitud, la frecuencia o la fase según la moduladora. Atendiendo a estas tres posibilidades se distinguen las modulaciones PAM, PWM y PPM, respectivamente (figura 21.25):

- PAM (*Pulse Amplitud Modulation*): la amplitud de los pulsos de la señal modulada varía de forma proporcional a la amplitud de la moduladora. La duración de los pulsos es constante. Se utiliza, fundamentalmente, como paso previo a la modulación PCM que se tratará posteriormente.
- PWM (*Pulse Width Modulation*): la duración de los pulsos varía de forma proporcional a la amplitud de la moduladora en el instante de muestreo. Este tipo de modulación se denomina también PDM (*Pulse Duration Modulation*). Se utiliza en convertidores digital a analógico, control de motores, control de convertidores electrónicos, etc. La modulación se realiza comparando la señal moduladora con

una señal en forma triangular en forma de diente de sierra, mientras en la demodulación se realiza simplemente con un filtro paso-bajo.

- PPM (*Pulse Position Modulation*): la posición de los pulsos varía de forma proporcional al valor de señal moduladora en el instante de muestreo.

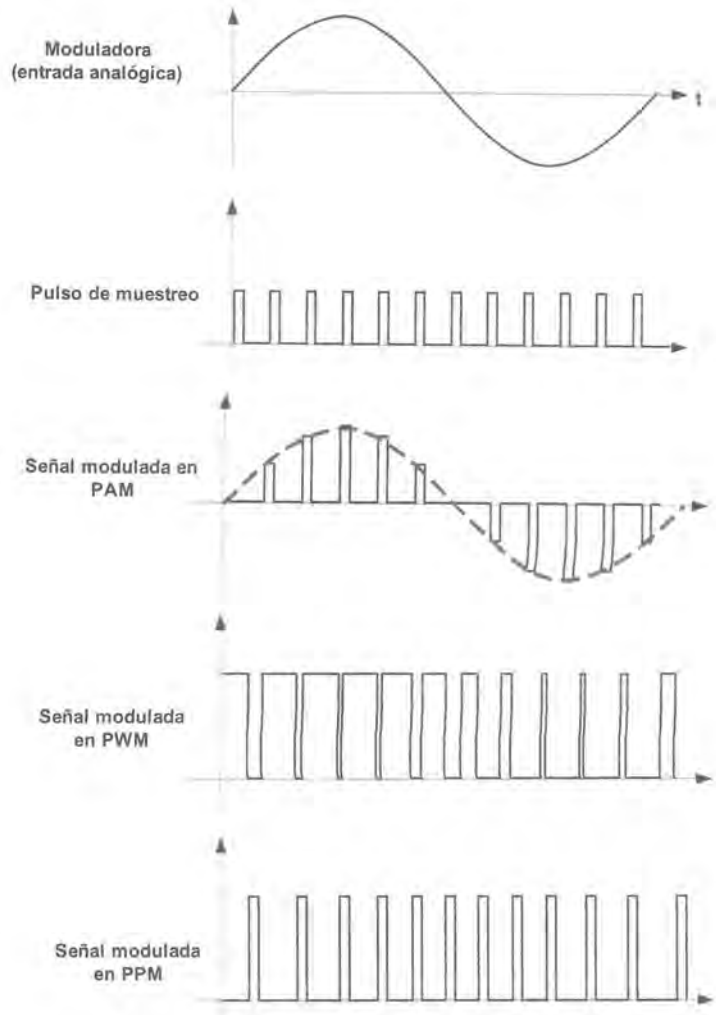


Figura 21.25. Modulación PAM, PWM y PPM.

De todas formas, el interés de estas modulaciones en la transmisión de señales es limitado. El método más utilizado es el PCM (*Pulse Code Modulation*). En este método se procede, en un primer momento, a una modulación PAM; posteriormente se cuantifica la amplitud del pulso y se codifica en un código binario (figura 21.26a). El elemento para llevar a cabo estas dos últimas acciones es el convertidor analógico a digital, que se estudiará en detalle en el capítulo 23. La demodulación se realiza con un convertidor digital a analógico que convierte la señal a PAM; la señal se recupera, finalmente, con un filtro paso bajo (figura 21.26b).

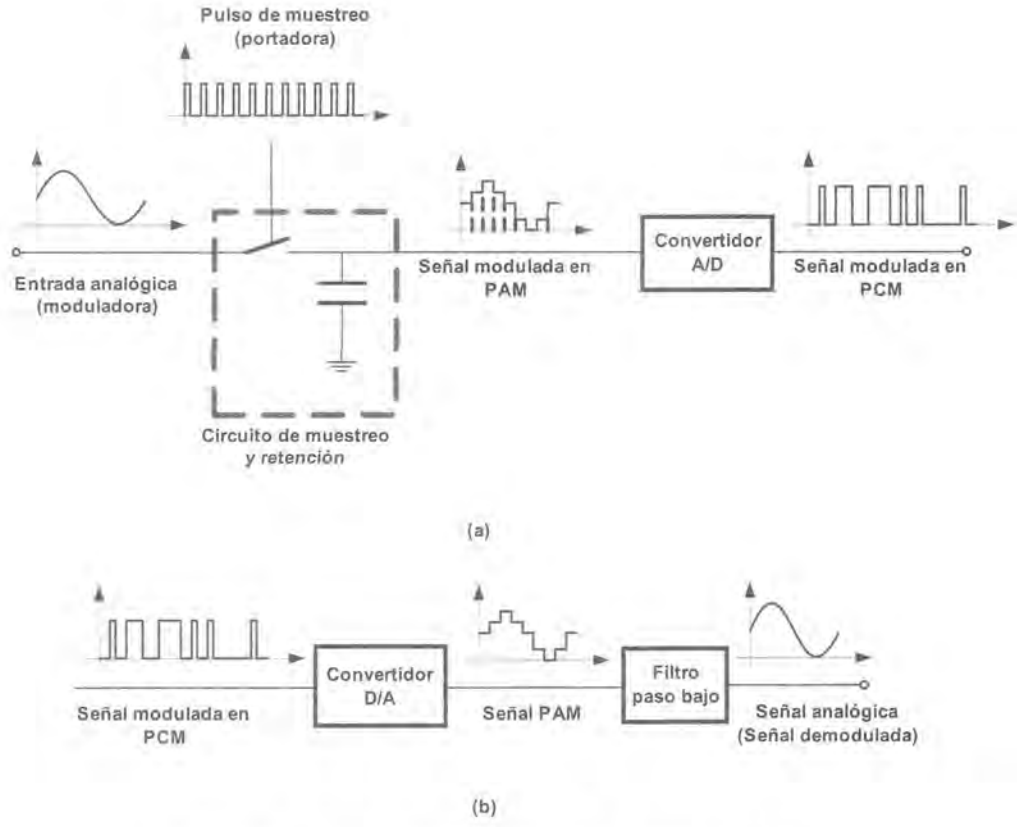


Figura 21.26. (a) Modulación PCM; (b) demodulación PCM.

Cuando se estudiaron las modulaciones AM y la ASK se demostró que el espectro de la señal modulada estaba comprendido, esencialmente, entre $f_p - f_m$ y $f_p + f_m$, donde f_p era la frecuencia de la portadora y f_m la frecuencia de la moduladora. En el caso de la modulación PAM (por supuesto, también en la PCM) hay que tener en cuenta que la portadora es una señal cuadrada por lo que su espectro está compuesto de señales de frecuencia 0, f_p, $2f_p$, $3f_p$..., por lo tanto, el espectro de la señal modulada se compone de bandas laterales alrededor de esas frecuencias (figura 21.27a). Si se solaparan las bandas laterales, se produciría la mezcla de

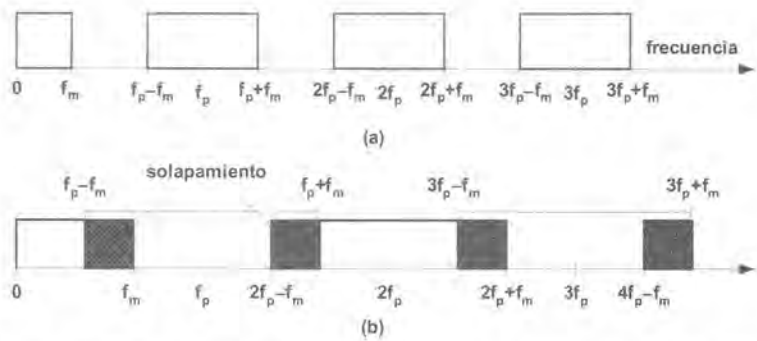


Figura 21.27. (a) Espectro de una señal modulada en PAM cuando la frecuencia de muestreo es más del doble de la frecuencia de la señal muestreada; (b) *aliasing*.

la información (*aliasing*) (figura 21.27b). Por ello, se debe garantizar que la frecuencia de muestreo (la frecuencia de la portadora) sea al menos dos veces la frecuencia de la señal analógica a muestrear (la moduladora). Normalmente, esto se garantiza limitando el propio ancho de banda de la señal analógica mediante filtros analógicos (filtros *antialiasing*).

Resumen

La información que se desea transmitir es la señal moduladora y se envía en forma de una modificación de la amplitud, la frecuencia o la fase de otra señal denominada portadora.

Se distinguen varios tipos de modulación atendiendo a si la señal portadora y la señal moduladora son analógicas o digitales.

Las modulaciones con portadora analógica y moduladora analógica más usadas son la AM y la FM.

En la modulación AM convencional el espectro de la señal modulada ocupa un ancho de banda igual a dos veces el de la señal moduladora.

En la modulación FM el ancho de banda ocupado es más grande que en AM y depende de las características concretas de la modulación.

Las modulaciones ASK, FSK y PSK son casos particulares de las modulaciones AM, FM y PM.

Para disminuir el ancho de banda ocupado en la transmisión, a veces se emplean modulaciones por portadora analógica y moduladora digital de orden superior a dos, como la QPSK.

Algunas modulaciones con portadora digital y moduladora analógica son la PAM, la PWM y la PPM.

Un tipo de modulación con portadora digital muy utilizada es la PCM que corresponde a una modulación PAM en la que, posteriormente, se cuantifica la amplitud del pulso y se codifica en un código binario.

Problemas propuestos

21.1 ¿Cuál es la frecuencia instantánea de la siguiente señal?

$$F(t) = 5\text{sen}\left(50000t + 10000t^2\right)$$

21.2 La siguiente función corresponde a una señal modulada en FM.

$$F(t) = 5\cos\left(10000t + 0{,}5\text{sen}\left(100t\right)\right)$$

¿Cuál es la frecuencia de la portadora? ¿Cuál es el índice de modulación y la desviación en frecuencia?

21.3 Se desea modular una señal de 1 Mbps en QAM y la portadora es de 10 MHz, ¿cuál es el ancho de banda ocupado?

21.4 Se desea transmitir por un mismo cable dos señales de 300 bps como máximo moduladas en FSK. A la primera señal se le asigna al cero una frecuencia de 1070 Hz y al uno 1270 Hz, a la segunda señal se le asignan las frecuencias 2025 Hz y 2225 Hz, respectivamente. ¿Cuál es el espectro de salida?

21.5 Se dispone de un transductor para medir una determinada variable que envía el resultado de la medida a un sistema de control en 4-20 mA. Sobre el mismo sistema de transmisión se desea enviar superpuesta una señal modulada en FSK de 1200 bps (de 5 mA de amplitud), para configurar diferentes parámetros del transductor desde el sistema de control. Se le asignará al cero la frecuencia de 2200 Hz y al uno, 1200 Hz. ¿Sería adecuada la solución planteada si el ancho de banda necesario para medir la variable es de 800 Hz? ¿Y si fuera de 10 Hz?

21.6 Una señal analógica proporcionada por el circuito de acondicionamiento de un determinado sensor se muestrea 100 Hz y se modula en PCM con un código de 8 bits. Se pretende enviar por radio esta señal modulada, a su vez, en ASK con una portadora de 433 MHz. ¿Cuál será el ancho de banda ocupado?

22

Variables muestreadas

22.1 Introducción
22.2 Variables analógicas, digitales y muestreadas
22.3 Teorema del muestreo
22.4 Variables submuestreadas
22.5 Reconstrucción de señales a partir de muestras
22.6 Muestreo y retención
22.7 Multiplexación de señales analógicas

22.1 Introducción

La utilización de procesadores digitales en los sistemas de medición actuales ha hecho posible obtener sistemas de gran fiabilidad, precisión y exactitud con el concurso de un conjunto de circuitos integrados digitales de muy bajo coste y consumo de energía y de una gran capacidad de almacenamiento de información. Pero esta última aún no es ilimitada por lo que resulta imposible manejar todo el conjunto de valores posibles de las variables analógicas de entrada.

Los datos que utiliza un procesador digital son valores discretos codificados que representan valores instantáneos de las señales continuas provenientes de los sensores y sus sistemas de acondicionamiento por lo que resultan necesarias interfaces que realicen la función de "atrapar" el valor de la señal en un instante dado y convertirlo en un código interpretable por el procesador.

También es necesario conocer con qué cantidad mínima de estos datos o muestras de la señal continua se puede establecer una relación de identidad entre la señal muestreada y la continua para la exactitud requerida de la medida, problema de gran importancia en el diseño de cualquier sistema de medida con el concurso de procesadores digitales.

22.2 Variables analógicas, digitales y muestreadas

En el capítulo 1 se hizo una primera definición sobre señales y variables. Una variable es:

- **Analógica** cuando los datos constituyen matemáticamente un conjunto denso, es decir, que puede tener cualquier valor dentro de un intervalo lo que implica que el número total de posibles valores es infinito.

- **Digital** cuando los datos constituyen un conjunto finito de valores; un caso particular de este tipo de variables es el sistema binario que permite dos valores diferentes denotados normalmente por los símbolos "0" y "1".

Cuando se realiza el procesamiento de variables analógicas para convertirlas en un formato digital, no resulta posible para los sistemas reales (además de que pudiera resultar innecesario) tener una representación digital de todos los infinitos valores de la variable analógica.

Una variable **muestreada** está formada por un conjunto de valores de una variable analógica espaciados regularmente en el tiempo. Si, además, esos valores pertenecen a un conjunto finito, la variable está **digitalizada**.

Para que la variable muestreada guarde una relación de identidad con la variable analógica que representa, la cantidad de muestras de esta última que conforma a la muestreada tiene que ser representativa, lo que implica que el número de valores de la muestreada debe ser elevado. Se sabe que en la medida que el número de muestras que deseamos tomar es mayor, mayor tiene que ser la frecuencia a la que se realice el muestreo. ¿Cuán grande debe ser esta frecuencia si a partir de la variable muestreada queremos reconstruir la señal analógica original?, ¿qué procedimientos se realizan para lograr el muestreo y la reconstrucción de señales a partir de muestras? A estas y otras interrogantes daremos respuesta en este capítulo.

22.3 Teorema del muestreo

Para obtener una variable muestreada a partir de una variable analógica se necesita producir una representación discreta temporal de la variable analógica, es decir, **tomar muestras** o **muestrear** a la variable analógica.

El **muestreo** (figura 22.1) es un procedimiento a través del cual, a partir de una variable analógica continua en el tiempo f(t), se obtiene una variable muestreada g(t) = f(kT) en tiempo discreto, donde k es un entero y T es el periodo de la señal de muestreo. Analíticamente, puede expresarse que:

$$g(t) = f(t)\big|_{t=k\cdot T} = f(kT)\text{ , donde } k = 0, 1, 2, \ldots. \qquad (22.1)$$

Figura 22.1. Procedimiento de muestreo.

El sistema que realiza este procedimiento recibe el nombre de **muestreador** y suele representarse como un simple interruptor ideal que cada T segundos hace que g(t) = f(t) y el resto del tiempo g(t) = 0, con lo que se consigue que g(t) = f(kT). De lo anterior se deduce que muestrear no es equivalente a mantener el valor de la muestra durante algún tiempo mas allá del instante de muestreo kT.

Si se utiliza una serie de impulsos de Dirac[1] h(t) como función de muestreo (figura 22.2b) y se realiza el producto de esta función por la variable analógica continua en el tiempo f(t) (figura 22.2a), se obtendrá la

[1] Esta función es un impulso de ancho infinitamente pequeño y amplitud unitaria, por lo que al realizar el producto de una función continua en el tiempo por una serie de impulsos de Dirac espaciados T segundos, se obtienen los valores de la función continua en los instantes T, 2T, 3T...kT.

variable muestreada g(t) (figura 22.2c) que no es más que un tren de impulsos modulados en amplitud por la señal analógica de entrada.

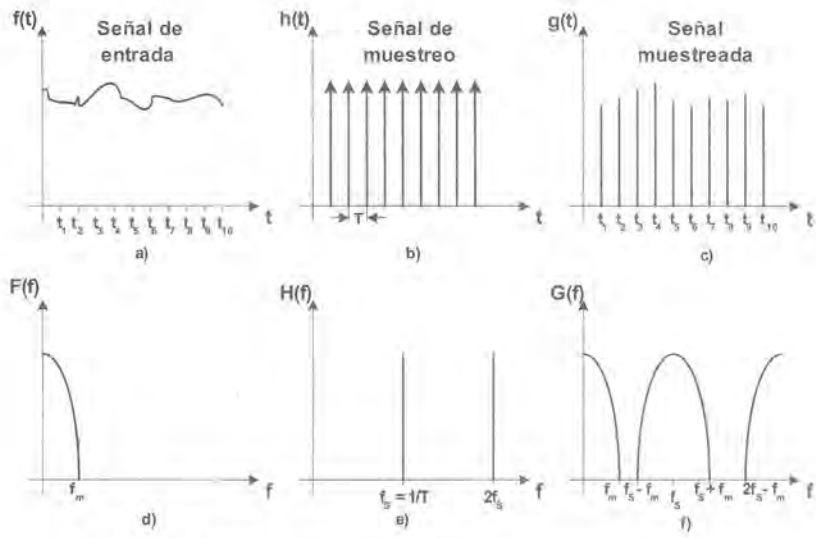


Figura 22.2. Ilustración del Teorema del muestreo.

Puede observarse que hemos obtenido un conjunto finito de valores de la variable analógica, espaciados regularmente por el periodo de la señal de muestreo ideal, a cuyo inverso, en adelante, llamaremos **frecuencia de muestreo** (f_s).

Es importante resaltar que la señal analógica f(t) es una señal de ancho de banda finito, lo que puede apreciarse en su espectro frecuencial (figura 22.2d) obtenido mediante la transformada de Fourier F(f) de la señal f(t). A su vez, el espectro frecuencial del tren de impulsos de muestreo (figura 22.2e) es una serie de componentes discretas de frecuencia múltiplos enteros de la frecuencia de muestreo f_s.

El muestreo realiza la convolución[2][22.6] de los espectros frecuenciales de la señal de entrada y la de muestreo, obteniéndose como resultado el espectro frecuencial de la señal muestreada (figura 22.2f). En él puede apreciarse la existencia (adicional al espectro frecuencial de la señal de entrada) de bandas laterales centradas en las frecuencias correspondientes a los múltiplos enteros de la frecuencia de muestreo, que son producidas por la modulación de amplitud[3] que tiene lugar.

En el caso representado en la figura 22.2 se ha escogido una frecuencia de muestreo f_s mayor que el doble de la máxima frecuencia del espectro de la señal ($f_s > 2f_m$). Pero, ¿qué obtendríamos si seleccionáramos $f_s < 2f_m$? La respuesta es obvia: se perderá información al producirse interferencia por solapamiento de parte del espectro de la señal y las componentes espectrales resultado de la modulación. La figura 22.3 presenta el espectro de la señal muestreada para este caso para frecuencias de hasta f_s.

El Teorema del muestreo de Nyquist afirma [22.1] que si f(t) es una señal de ancho de banda finito, esta puede ser reconstruida en cualquier punto a partir de sus valores muestreados f(kT) si el muestreo se ha efectuado a una frecuencia f_s que, al menos, sea el doble de la máxima frecuencia del espectro de la señal, o sea que

$$f_s \geq 2f_m \tag{22.2}$$

[2] Existe una equivalencia entre la multiplicación en el espacio temporal y la convolución en el espacio frecuencial. El lector puede consultar la literatura especializada en esta materia.

[3] En el capítulo 21 el lector puede encontrar los fundamentos de la modulación de amplitud.

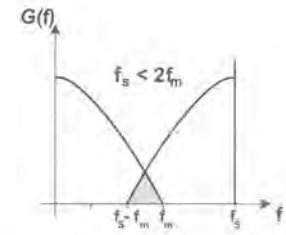


Figura 22.3. Solapamiento de los espectros para $f_s < 2f_m$.

A la condición "igual a" de la expresión (22.2) se conoce con el nombre de frecuencia de Nyquist [22.2] y si el muestreo es realizado a una frecuencia mayor que la de Nyquist, al menos en teoría y bajo las condiciones anteriormente expuestas, no ocurrirá solapamiento y toda la información de la señal puede ser extraída del espectro de la señal muestreada mediante eliminación o filtrado de las componentes de la modulación.

Si la señal analógica continua f(t) tiene transformada de Fourier continua $F(\omega)$, se cumplirá que:

$$F(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(t)\cdot e^{-j\omega t}dt \tag{22.3}$$

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)\cdot e^{j\omega t}d\omega \tag{22.4}$$

La ecuación (22.1) ofrece la relación existente entre f(t) y g(t) por lo que (22.4) puede expresarse para g(t) mediante:

$$g(t) = f(kT) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)\cdot e^{j\omega kT}d\omega \tag{22.5}$$

Si esta señal muestreada posee transformada discreta de Fourier $F_d(\omega)$, también estará relacionada con su transformada discreta mediante:

$$f(kT) = \frac{1}{2\pi}\int_{-\pi}^{\pi} F_d(\omega)\cdot e^{j\omega k}d\omega \tag{22.6}$$

y puede demostrarse [22.2] [22.4] que la transformada discreta de Fourier de la señal muestreada y la transformada continua guardan la siguiente relación:

$$F_d(\omega) = \sum_{p=-\infty}^{\infty} F(\omega + 2\pi f_s p) \quad \text{para } p = 0, \pm 1, \pm 2, \ldots \tag{22.7}$$

Esto significa que si la señal analógica tiene un espectro como el representado en la figura 22.2d expresado analíticamente por su transformada continua de Fourier $F(\omega)$ (en esta figura sólo se ha presentado la parte real de los espectros pero las expresiones anteriores incluyen la parte imaginaria), el espectro de la señal muestreada f(kT) será una repetición periódica del espectro de f(t) cada $p\omega_s$, donde $\omega_s = 2\pi f_s$ como se muestra en la figura 22.2f y, analíticamente, en la expresión (22.7).

Para que no exista solapamiento (figura 22.3) y con ello se pueda reconstruir f(t) a partir de sus muestras f(kT) sin pérdida de información, tiene que cumplirse que el espectro de la señal analógica sea limitado en frecuencias hasta $\omega_m = 2\pi f_m$ y que la frecuencia de muestreo $f_s \geq 2f_m$.

Ahora bien, los procesos de muestreo reales son bien diferentes. En primer lugar, la señal de muestreo no suele ser un tren de impulsos, sino de pulsos de muy corta duración δ y de amplitud A (figura 22.4b), con lo que su espectro frecuencial cambia sustancialmente debido a que, adicionalmente a las componentes fre-

cuenciales corrrespondientes al impulso, el pulso tiene otras componentes. La figura 22.4e presenta la envolvente de la transformada de Fourier correspondiente a esas componentes [22.2] cuya expresión es:

$$E = \frac{A\delta}{T} \cdot \frac{\text{sen}(\pi f \delta)}{\pi f \delta} \tag{22.8}$$

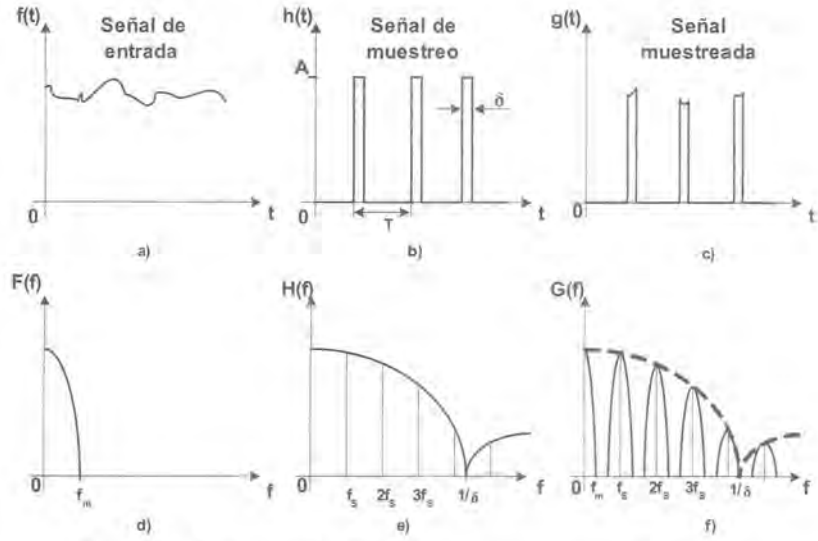


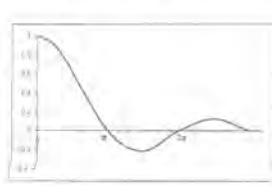
Figura 22.4. *Ilustración del muestreo con un tren de pulsos.*

Obsérvese que posee ceros en las frecuencias múltiplos de $1/\delta$ y que la distorsión de amplitud del espectro provocado por la envolvente del espectro del tren de pulsos (función $\text{sinc}(x)$[4]) es menor en la medida que lo es δ. Al producirse la convolución de los espectros de esta señal y de la señal analógica, el resultado es el mismo que si se utiliza un tren de impulsos de Dirac, aunque con la ponderación correspondiente a la envolvente del espectro frecuencial del pulso.

La figura 22.4f presenta de forma aproximada el espectro de la variable muestreada donde puede apreciarse que el efecto de ponderación sobre el espectro de la señal en buena medida ayuda al filtrado de las componentes de la modulación ya que atenúa pronunciadamente las componentes de la modulación de altas frecuencias por lo que es ventajoso utilizar un tren de pulsos como señal de muestreo desde este punto de vista.

La atenuación en el espectro de la señal puede minimizarse si el ancho del pulso δ es muy pequeño en comparación con el periodo T. Un caso muy interesante se presenta cuando el valor de la muestra en el instante inicial del pulso es mantenido en la salida del muestreador hasta el instante en que se inicie la muestra siguiente, es decir, cuando $\delta = T$. La figura 22.5a ilustra la señal analógica y la señal muestreada como una función del tiempo y la figura 22.5b el espectro frecuencial de la señal muestreada.

[4] La función $\text{sinc}(t) = \text{sen}\pi t/\pi t$ es la respuesta al impulso de un filtro paso bajo ideal con frecuencia de corte en la mitad de la frecuencia de repetición de los impulsos.



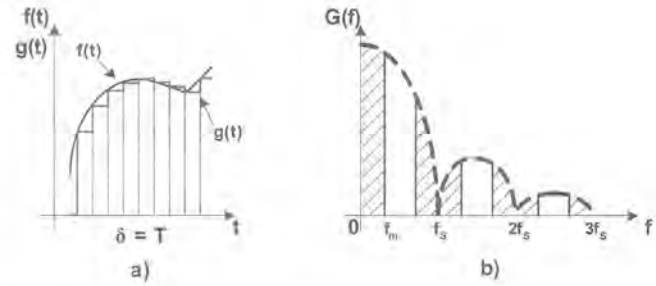


Figura 22.5. Espectro de la señal muestreada con pulsos de $\delta = T$.

Obsérvese que en estas condiciones la reducción de las componentes de alta frecuencia es óptima, pero se requiere compensar de la deformación que sufre el espectro de amplitud de la señal. Esta compensación puede lograrse aumentando la ganancia del filtro paso bajo (que limita el espectro de la señal) en las frecuencias cercanas a f_m o digitalmente en el procesador digital al cual se entregan los valores de la señal muestreada después de cuantificados y codificados por un convertidor analógico-digital.

Finalmente, si se considera que la señal no tiene ancho de banda finito, entonces existe la probabilidad de que un conjunto de componentes espectrales de la señal tengan frecuencias superiores a la mitad de la frecuencia de muestreo y esto producirá el solapamiento incontrolado del extremo superior del espectro de la señal con la banda lateral de la modulación de amplitud que se representa en la figura 22.3, con la pérdida de información asociada.

22.3.1 EL PROBLEMA DEL SOLAPAMIENTO O *ALIASING*

Hemos visto que el solapamiento o *aliasing* es una distorsión que se produce en el proceso de muestreo siempre que $f_s < 2f_m$, pero queremos resaltar la importancia práctica de esta distorsión.

Supongamos que f(t) sea una señal analógica senoidal de frecuencia f_a que es muestreada con una frecuencia f_s. Si $f(t) = A\cos(2\pi f_a t + \varphi_a)$, de la expresión (22.1) se conoce que la señal muestreada g(kT), también será senoidal en tiempo discreto y su expresión será:

$$g(kT) = A\cos(2\pi k f_d + \varphi_a)$$

donde $f_d = f_a/f_s$ es la frecuencia discreta de g(kT).

La expresión anterior nos demuestra que podemos obtener señales muestreadas idénticas a partir de señales de igual fase y amplitud pero de distintas frecuencias siempre que la frecuencia discreta f_d sea la misma, lo que se consigue con diferentes frecuencias de muestreo. Esta incertidumbre en el muestreo es la esencia del *aliasing*.

Desde el punto de vista de la reconstrucción de la señal a partir de sus muestras, esta incertidumbre ofrece la siguiente solución. Supongamos que tenemos dos señales senoidales, una de frecuencia igual a 200 Hz y otra de 1000 Hz, y que realizamos el muestreo a una frecuencia $f_s = 800$ Hz. La expresión de g(kT) para ambas señales **es la misma**:

$$g_1(kT) = \mathbf{A\cos(0{,}5\pi k)} \text{ y } g_2(kT) = A\cos(2.5\,\pi k) = A\cos(2.5\,\pi k - 2\pi k) = \mathbf{A\cos(0{,}5\pi k)}$$

Si estas secuencias se hacen llegar a un convertidor digital-analógico para que ofrezca a su salida una señal analógica, la función reconstruida responderá a la señal de frecuencia 200 Hz, para la cual el criterio de Nyquist se cumple y **no es posible** recuperar la señal de 1000 Hz a partir de sus muestras.

Este fenómeno se puede apreciar en las películas en las que al observar una rueda que gira "nos parece" que gira a una velocidad angular mucho menor que la real. Ello se debe a que la película es una secuencia de fotos (donde cada foto es una muestra), por lo general de frecuencia 24 fotos/s, y la rueda gira a una velocidad angular mayor que la mitad de esta frecuencia.

Para evitar estos efectos del solapamiento resulta imprescindible la utilización de un filtro *anti-aliasing*, cuyas características vincularemos con la selección de la frecuencia de muestreo. El diseño de filtros analógicos ha sido tratado en detalle en el capítulo 4, por lo que aquí sólo apuntaremos algunas características específicas de sistemas integrados para este propósito.

22.3.2 SELECCIÓN DE LA FRECUENCIA DE MUESTREO

Hasta ahora hemos podido comprobar que para poder recuperar la señal de entrada a partir de sus muestras, hay que limitar (si no lo es) el espectro de frecuencias de la señal analógica de origen hasta una cierta frecuencia máxima f_m y la frecuencia de muestreo debe ser mayor que el doble de esta frecuencia.

Para seleccionar la frecuencia de muestreo de las señales analógicas de espectro conocido puede aplicarse directamente el criterio de Nyquist con las consideraciones que analizaremos más adelante en el apartado 22.3.3.

Si no se conoce el espectro de la señal analógica f(t) pero se tiene un histograma estadísticamente valioso de sus valores medidos a una frecuencia de muestreo muy alta, podrá encontrarse la mayor rapidez de cambio de esta señal con el tiempo evaluando su pendiente máxima. La figura 22.6 representa la región de la señal analógica f(t) en la que se produce su mayor rapidez de cambio con el tiempo.

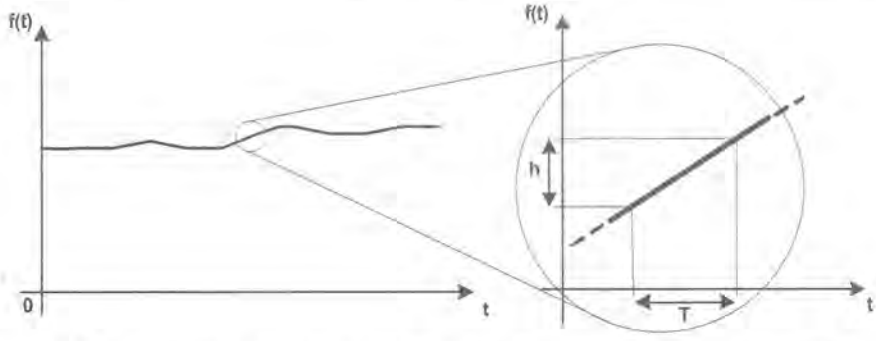


Figura 22.6. Selección de f_s para una señal de espectro frecuencial no conocido.

Si en esta región se hace que h sea la resolución[5] del convertidor A/D que sigue al muestreador, puede seleccionarse $f_s = 1/T$ y se tendrá la garantía de tomar información contenida en las componentes de más alta frecuencia de f(t) con la resolución del sistema. Este es un criterio que cumple muy sobradamente el Teorema del muestreo lo que puede demostrarse si se aplica a una señal de espectro conocido (véase el Ejercicio 23.9) por lo que sus resultados no siempre ofrecen la mejor solución económica a este problema ya que imponen requisitos dinámicos muy exigentes al convertidor A/D, pero demuestran que para hacer una reconstrucción de la señal analógica original con la fidelidad correspondiente a la resolución del sistema la frecuencia de muestreo debe ser muchas veces mayor que la de Nyquist (sobre-muestreo u *oversampling*) si

[5] La resolución de un convertidor A/D es: resolución = FS/(N-1) = FS/(2^n-1), donde FS es el fondo de escala de la entrada analógica, N el número de estados de la salida (que generalmente se expresa en una potencia de dos, $N = 2^n$, donde n es el número de bits de la palabra binaria que representa a cada estado). Para más detalle consulte el capítulo 23.

se emplea un bloqueador de orden cero[6] [22.5] como elemento generador de la señal analógica a partir de sus muestras.

EJERCICIO 22.1

Un proceso transitorio produce una señal en el margen de 0 ÷ 5 V cuya máxima rapidez de cambio es de 3800 V/s. Se desea muestrear de manera tal que la tensión entre muestras consecutivas no exceda nunca de 10 mV, que es la resolución del convertidor A/D del sistema. ¿Cuál es la mínima frecuencia de muestreo que satisface estas condiciones?

SOLUCIÓN:

La mínima frecuencia de muestreo será aquella que iguale a la máxima rapidez de cambio de la señal, es decir, $f_{smin} = 3800/10 \cdot 10^{-3} = 380$ kHz.

La frecuencia de muestreo del convertidor A/D que se utilice deberá ser igual o mayor que esta frecuencia mínima.

La selección de un valor para la frecuencia de muestreo f_s estará determinada no sólo por la frecuencia de Nyquist; también dependerá de otros factores relacionados con las limitaciones prácticas que tienen los sistemas reales que intervienen en la conversión de la señal analógica en una variable muestreada, cuantificada y codificada que de inmediato analizaremos.

22.3.3 LIMITACIONES

La selección de la frecuencia de muestreo tiene un límite inferior que no sólo depende de la frecuencia de Nyquist. El filtro paso bajo que limita el espectro frecuencial de la señal analógica o filtro *anti-aliasing* no tiene una característica ideal, es decir, que si su frecuencia de corte se selecciona en la frecuencia f_m la frecuencia de muestreo mínima a seleccionar (f_{smin}) tendrá necesariamente que ser mayor que $2f_m$ si aplicáramos estrictamente el Teorema del muestreo.

Sin embargo, como la resolución del sistema de conversión A/D que sigue al muestreador no es cero, podemos permitirnos algún solapamiento (figura 22.7). Para que el solapamiento de la parte sombreada del espectro no afecte a la señal, el filtro tendrá que ofrecer una atenuación A a la frecuencia $f_s - f_m$ tal que la amplitud de la señal de esa frecuencia sea menor que la resolución del sistema de cuantificación que sucede al muestreador.

Si suponemos que todas las componentes espectrales de f(t) pueden tener amplitud máxima igual al margen estático de la señal (que es el peor de los casos) y si la resolución del convertidor A/D al cual el muestreador le entregará las muestras de la señal analógica es la correspondiente a n bits, se requiere una atenuación de A dB (por ejemplo, A = 48 dB para n = 8 bits) para que la amplitud de la componente espectral de frecuencia $f_s - f_m$ sea menor que el peso del LSB[7] del convertidor A/D, donde A vendrá dada por:

$$A = 20\log \text{resolución}_N \qquad (22.9)$$

---

[6] Un bloqueador de orden cero mantiene el valor digital de la muestra durante el periodo de muestreo T. Es la operación típica de un convertidor D/A.

[7] En un convertidor A/D el peso de 1 LSB (**L**east Significant **B**it) es su **resolución**. Si FS es el fondo de escala de la entrada analógica, N el número de estados (que se expresa en una potencia de dos, $N = 2^n$, donde n es el número de bits de la palabra binaria que representa a cada estado), entonces la resolución será: resolución = 1 LSB = $FS/(2^n-1)$. Esto se verá en más detalle en el capítulo 23.

Donde resolución$_N$[8] es la resolución normalizada del convertidor A/D que sigue al muestreador.

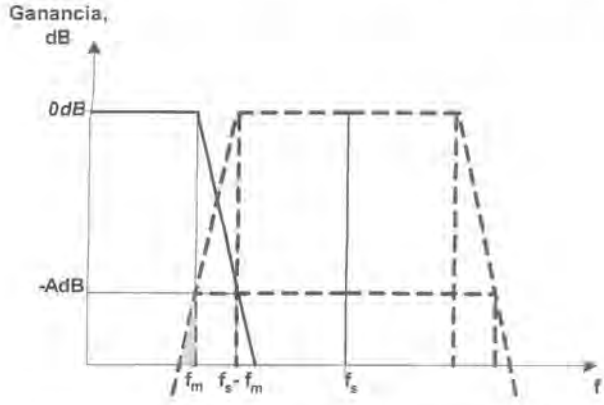


Figura 22.7. Ilustración del límite mínimo de f_s.

EJERCICIO 22.2

Se desea limitar el espectro de una señal analógica unipolar hasta la frecuencia de 1000 Hz para ser muestreada a una frecuencia f_s = 5000 Hz. Si la resolución normalizada del convertidor A/D que sucede al muestreador es 0,003906 (correspondiente a 8 bits), calcule la pendiente de atenuación del filtro anti-aliasing necesaria para evitar los efectos del solapamiento en el espectro apreciado de la señal.

SOLUCIÓN:

Si suponemos el peor de los casos en que a la frecuencia $f_s - f_m = 4000$ Hz la amplitud de la señal puede ser la máxima (margen estático del convertidor A/D), el filtro a utilizar a esta frecuencia debe ofrecer la atenuación dada por (22.9):

$$A_{4000Hz} = 20\log 0{,}003906 = -48 \text{ dB}$$

Esta frecuencia se corresponde con el cuádruplo de la frecuencia f_m, es decir, dos octavas por encima de ella, por lo que la atenuación del filtro tendrá que ser de −24 dB/oct y su frecuencia de corte seleccionada en $f_c = f_m = 1000$ Hz.

Una solución pudiera ser el empleo de un filtro de Butterworth, 4º orden, pero pueden evaluarse otras variantes. Se sugiere el repaso del capítulo 4 para abordar este diseño.

El ejemplo anterior es un caso extremo ya que, por lo general, si se limita el espectro de una señal hasta una frecuencia f_m es porque a partir de esa frecuencia el contenido de información del espectro puede despreciarse, entre otras cosas, porque las amplitudes de las componentes espectrales resultarán muy pequeñas en comparación con las del espectro apreciado por lo que no siempre se requiere una característica de atenuación de pendiente tan pronunciada para el filtro; no obstante, el coste y la complejidad del filtro dependen directamente de lo próximas que estén f_m y $f_s - f_m$.

[8] La resolución normalizada es la resolución dividida por el fondo de escala de la entrada. Analíticamente: resolución$_N$ = resolución/FS.

Pero si la limitación se realiza porque no interesa la información contenida en esa parte superior del espectro, entonces el criterio tomado para la selección de la atenuación del filtro en el ejemplo anterior es adecuado.

Filtros *anti-aliasing* integrados

Conseguir grandes atenuaciones en un margen de frecuencias relativamente pequeño es la tarea que se encomienda a un filtro *anti-aliasing*. Ello implica la utilización de filtros de tercer, cuarto u órdenes superiores, siendo los que responden a la aproximación de Butterworth los más comúnmente utilizados ya que ofrecen una respuesta plana en la banda de paso (véase el capítulo 4). Si se construye el filtro con componentes discretos, el coste será alto y ocuparán bastante espacio en los circuitos impresos. Es por ello que se prefiere utilizar sistemas integrados que ofrecen una solución técnico-económica más ventajosa.

En algunos sistemas comerciales, la frecuencia de corte del filtro ($f_c = f_m$) depende de la frecuencia y la estabilidad de un reloj externo f_{CLK}, que guarda una relación constante con la frecuencia de corte. Por ejemplo, el filtro *anti-aliasing* de capacidad conmutada TLC04 de Texas Instruments, cuyas hojas de características se ofrecen en el CD, puede conseguir frecuencias de corte desde 0,1 Hz hasta 30 kHz, controlada por un reloj externo en relación 50:1 con la frecuencia de corte o por una resistencia R y un condensador C externos. Responde al algoritmo de Butterworth de cuarto orden [22.6].

Debido a estas consideraciones no resulta beneficioso que la frecuencia de muestreo **sea inferior a 5 f_m** si se pretende recuperar la información original a partir de las muestras y limitar el ancho de banda de la señal hasta f_m con un filtro, y un criterio práctico aconseja seleccionarla entre 7 y 20 veces mayor que f_m para no imponer exigencias muy altas al filtro.

El convertidor analógico-digital que realiza la conversión de la señal analógica muestreada a un código digital, está limitado en su máxima rapidez de conversión, lo que impone un **límite superior** práctico a la selección de la frecuencia de muestreo. Este límite está en función de la resolución con la que se necesite realizar la conversión y del tipo de convertidor que se utilice y, generalmente, su coste está en función directa del número de bits del código digital con que se representen los valores de la variable muestreada y de su máxima rapidez de conversión.

Por lo que la selección de la frecuencia de muestreo es un compromiso técnico-económico entre la calidad y coste del filtro analógico que limita el ancho de banda de la señal y del convertidor A/D que se seleccionen para el canal de muestreo, cuantificación y codificación de la señal analógica.

La figura 22.8 presenta la arquitectura general del sistema de medida con el concurso de un procesador digital.

Figura 22.8. Arquitectura general de un canal de muestreo y adquisición de una señal analógica.

Con línea de puntos se ha representado el sistema de muestreo al que hemos denominado de forma más completa de "muestreo y retención" ya que, como veremos más adelante, los circuitos prácticos que realizan esta función no sólo muestrean la señal, sino que también "retienen" el valor de la muestra durante un tiempo debido a que algunos convertidores A/D requieren que su entrada sea constante durante su tiempo de conversión. De esta forma, la señal de entrada es la señal analógica proveniente del sistema sensor-acondicionador, la salida del filtro *anti-aliasing* es la señal analógica limitada en ancho de banda f(t) y la entrada al convertidor A/D es la señal muestreada $g(t) = f(k/f_s)$.

22.4 Variables submuestreadas

Se dice que una variable es sub-muestreada cuando es el resultado de aplicar a la señal analógica una frecuencia de muestreo inferior que la frecuencia de Nyquist.

Este es un caso que pudiera parecer sin interés desde el punto de vista de recuperar la señal analógica continua a partir de sus muestras, por cuanto el *aliasing* incrementaría el error de aproximación en la misma medida que la relación f_s/f_m sea menor de dos y, de limitar el ancho de banda mediante un filtro *anti-aliasing*, las componentes espectrales filtradas impedirán que la señal recuperada a partir de las muestras tenga un pequeño error de aproximación con respecto a la señal original. Pero si lo que se desea es obtener una relación directa entre el valor de la señal y la muestra en un instante dado, sin la pretensión de reproducir la señal a partir de sus muestras, entonces este aspecto cobra particular importancia. En muchos sistemas de medida, incluso de monitorización de variables analógicas, se produce el muestreo con una frecuencia muy inferior a la de Nyquist sin emplear un filtro que limite el ancho de banda de la señal y que impida los efectos del *aliasing* y la incidencia de ruidos de altas frecuencias sobre la medida aproximada que se realiza.

22.5 Reconstrucción de señales a partir de muestras

Hemos visto que si la función muestreada f(kT) es una secuencia de valores de f(t) y se han cumplido las condiciones del teorema de Nyquist, en cualquier momento se puede reconstruir la señal f(t) a partir de los valores discretos de f(kT).

El sistema encargado de realizar la conversión del código digital en una señal analógica recibe el nombre de convertidor digital-analógico (véase capítulo 23), que es un sistema que genera en su salida una señal de corriente o tensión en forma de escalera de peldaños de duración $T = 1/f_s$ durante la cual el valor de la señal, idealmente, es el valor de la muestra correspondiente de la secuencia temporal f(kT), seguido de un filtro que elimina las componentes espectrales no deseadas. Realmente, el valor de la señal de salida de un convertidor D/A es el valor muestreado y cuantificado de la señal analógica por lo que no es en realidad una señal analógica, sino digital.

La figura 22.5a ilustra la forma de onda que se produce en la salida del convertidor D/A, allí representada como g(t). El espectro de esta señal es de la forma representada en la figura 22.5b, por lo que filtrando las componentes espectrales correspondientes a la modulación de amplitud que tuvo lugar, se puede obtener la señal f(t) original, al menos en teoría.

La figura 22.9 presenta la arquitectura general del sistema que reproduce la señal analógica a partir de sus muestras.

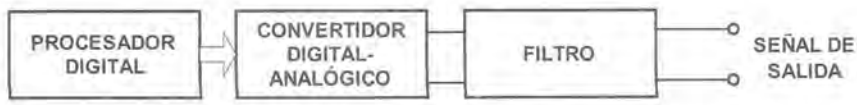


Figura 22.9. Sistema de reconstrucción de una señal analógica a partir de sus muestras.

Desde el punto de vista del filtrado analógico necesario para recuperar la señal analógica f(t) es conveniente que el periodo de muestreo T sea muy pequeño, con lo que las exigencias técnicas para el filtro paso bajo de la salida del convertidor digital-analógico serán menores. Pero los convertidores D/A reales imponen

una limitación a la duración mínima del periodo de muestreo T (equivalente al inverso de su rapidez de conversión) que depende de su resolución y del tipo de convertidor por lo que el proceso de reconstrucción de señales a partir de sus muestras también pasa por una solución de compromiso entre la complejidad y el coste del convertidor D/A y el filtro de salida.

Una forma de reconstruir señales a partir de muestras, disminuyendo el error de aproximación entre la señal original y la reconstruida, es mediante la aplicación de un procesamiento digital utilizando algoritmos de interpolación [22.6].

A partir de las muestras f(kT) se genera una nueva secuencia de muestras espaciadas T/N, donde N es mayor que uno. La figura 22.10 muestra un esquema de este procedimiento, conocido como *N upsampling*[9].

A partir de las muestras f(kT) (figura 22.10a) se interpolan nuevos valores y se conforma f(kT/N) (figura 22.10b), que se entrega al convertidor D/A que necesariamente tiene que tener una rapidez de conversión igual o superior a 1/NT, que es la "nueva" frecuencia de muestreo y, a partir de ellas, este genera f'(t) (figura 22.10c). Las figuras 22.10d y 22.10e presentan la relación que existe entre los espectros de f(kT) y de f(kT/N), cuyas envolventes son G(f) y G'(f), respectivamente. A los efectos del filtro, la recuperación del espectro de la señal f(t), que no cambia por el *upsampling*, resulta mucho menos exigente, pero se han aumentado las exigencias dinámicas del convertidor D/A e incorporado un procesamiento digital de la señal en el procesador.

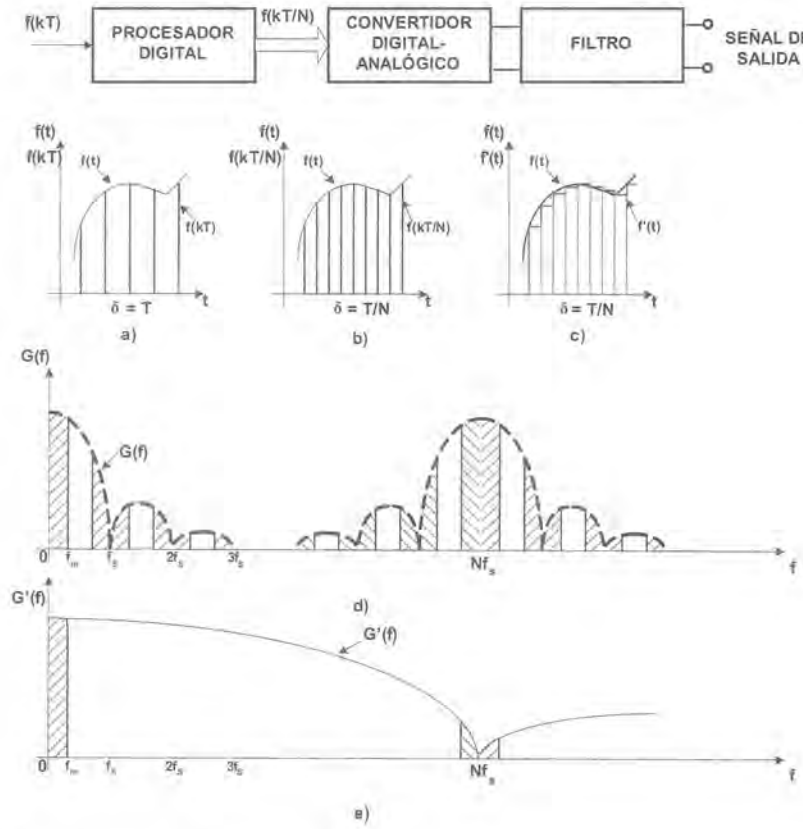


Figura 22.10. Procedimiento de *upsampling* para la recuperación de señales a partir de muestras.

[9] Aunque la traducción al español de este término es sobre-muestreo, usamos "*upsampling*" pues se refiere a un **procesamiento digital**. A diferencia, el sobre-muestreo (*oversampling*) es un **procesamiento analógico**: tomar muestras a una frecuencia mayor que $2f_m$.

22.6 Muestreo y retención

Un sistema de muestreo y retención o *sample-and-hold* (S/H o S&H) es un circuito que posee una entrada analógica, una entrada de control digital S/H y una salida analógica, como se muestra en la figura 22.11a.

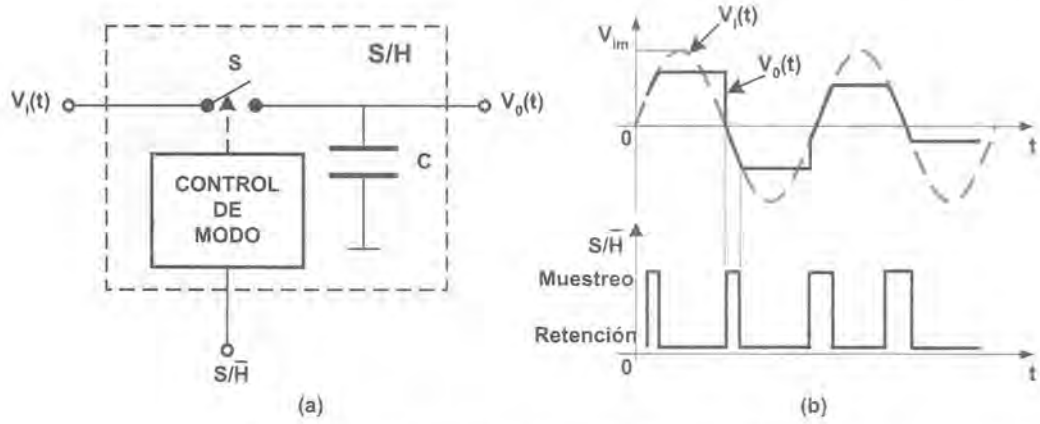


Figura 22.11. Sistema de muestreo y retención.

Sus elementos básicos son un interruptor S controlado por un circuito de control digital que, por lo general, hace que permanezca cerrado cuando la señal digital de entrada se encuentra en "1" (modo S, muestreo o *sample*), haciendo que $v_0(t) = v_i(t)$ (figura 22.11b) y abierto cuando la señal S/H se encuentra en "0" (modo H, retención o *hold*) y un condensador C que, idealmente, mantiene en la salida la tensión existente en la entrada en el momento del cambio del modo S al H hasta que se produzca un nuevo cambio de estado.

Al producirse el cambio de modo de retención a muestreo, la salida "salta" idealmente hasta atrapar el valor de la entrada y sigue a la entrada hasta que se produzca un nuevo cambio de modo.

En los circuitos de muestreo y retención reales estas características no se cumplen del todo debido a la existencia de un valor (aunque sea pequeño) de resistencia en el interruptor cerrado S y un valor muy alto de resistencia en el interruptor abierto que, afectados por las resistencias de la fuente de señal $v_i(t)$ y de la carga donde se desarrolla $v_0(t)$, hacen que el condensador no cambie con constante de tiempo nula su tensión durante el modo muestreo y que pierda carga durante el modo retención.

22.6.1 EL MODO MUESTREO

La figura 22.12 presenta el circuito equivalente del sistema de muestreo y retención en el modo muestreo y su respuesta a una entrada escalón de tensión. Esta situación es perfectamente posible ya que en el momento en que se pasa a este modo, la tensión anterior en el condensador puede diferir de la actual en el máximo valor del margen estático de la señal.

Si R_s representa la resistencia R_{ON} del interruptor en serie con la resistencia de la fuente de señal $v_i(t)$ y R_p representa la resistencia de la carga y es muy grande, se puede considerar que el circuito se comporta como un filtro paso bajo con constante de tiempo R_sC. En estas condiciones la salida no será una réplica aproximada de la entrada contenida dentro de una banda de error hasta que no haya transcurrido una duración t_A que se conoce con el nombre de **tiempo de adquisición** (*acquisition time*) del sistema S/H. Este es uno de los parámetros de singular importancia del circuito y suele especificarse con las condiciones de valor

de C y de margen del error relativo al margen de entrada (0,01%, 0,001%). Durante el tiempo que el circuito se mantenga en este modo, la salida "seguirá" a la entrada con una diferencia dependiente de su ancho de banda. Los circuitos en los que el "rastreo" de la entrada se realice con gran precisión reciben el nombre de *track-and-hold* (T/H) o *track-and-hold amplifier* (THA).

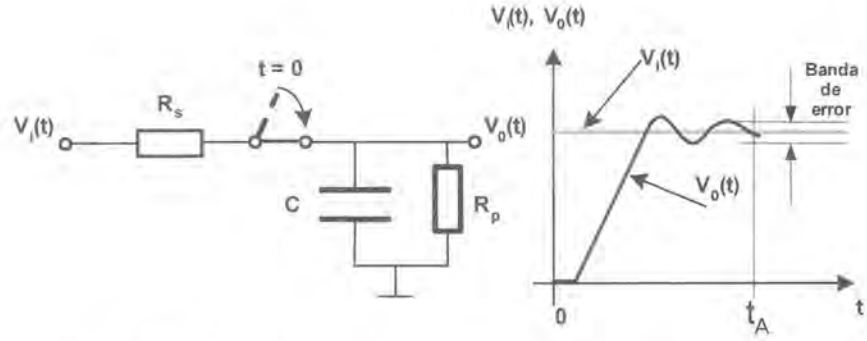


Figura 22.12. Ilustración del transitorio del modo muestreo.

22.6.2 EL MODO RETENCIÓN

La figura 22.13 presenta el circuito equivalente del sistema en el modo retención suponiendo que el interruptor S ha pasado de una resistencia R_{ON} muy pequeña a otra R_{OFF} muy alta de forma inmediata, donde R representa la resistencia equivalente paralela de la carga R_p, del condensador y la resistencia R_{OFF} del interruptor, y R_s es la resistencia interna de la fuente de señal $V_i(t)$.

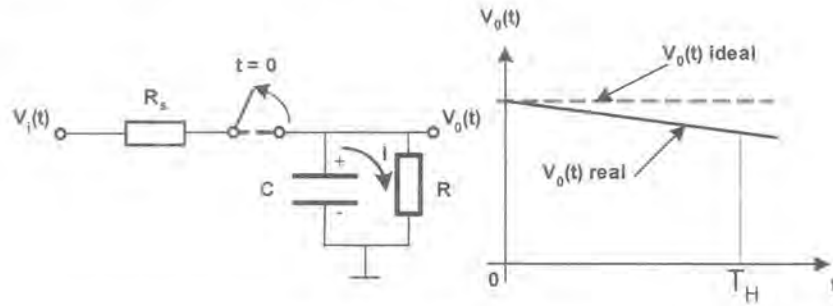


Figura 22.13. Ilustración de la salida durante el modo retención.

Si asumimos que la corriente de fugas del condensador I es constante durante tiempos de retención T_H pequeños en comparación con la constante de tiempo RC, la salida real caerá con una **pendiente** (*droop rate*) constante –I/C en voltios por segundo (V/s) y, al finalizar T_H, el decaimiento será de T_HI/C. Si el sistema de muestreo y retención entregara su salida a un convertidor A/D, el decaimiento durante su tiempo de conversión no puede ser mayor que la mitad de su resolución para evitar errores en la conversión.

Otro fenómeno presente en este modo y no representado en la figura 22.13 porque depende del valor, la polaridad y la frecuencia de la entrada es la presencia de la entrada en la salida producto de la no-idealidad del interruptor abierto y la capacidad entrada-salida. Para los circuitos prácticos se cuantifica como la **razón de atenuación del paso** de la señal a través del interruptor abierto (*feedthrough attenuation ratio*) y se expresa en decibelios para una frecuencia dada. Si el sistema de muestreo y retención entrega su salida a un

convertidor A/D, la atenuación al paso de la señal durante el modo H no podrá ser mayor que la mitad de su resolución.

Puede demostrarse que la expresión para la razón de atenuación necesaria (A_F) al paso de la señal para que la presencia de la entrada en la salida no exceda a la mitad de la resolución (½LSB) del convertidor A/D en función de su número de bits n es: $A_F \geq 6(n + 1)$.

EJERCICIO 22.3

Un convertidor A/D de 16 bits y margen estático 0 ÷ 10 V requiere que su señal de entrada permanezca constante durante 10 μs para realizar la cuantificación del valor adquirido de una señal analógica de margen estático 0 ÷ 10 V y limitada en su ancho de banda hasta 1 kHz. ¿Qué decaimiento máximo se le permite al S/H? ¿Cuál es la razón de atenuación mínima de la entrada en la salida?

SOLUCIÓN:

Si el convertidor A/D es de 16 bits y su margen de entrada es 10V, su resolución aproximada será: resolución = 1 LSB = 10 V/(2^{16}) = 152,59 μV y la mitad será ½LSB = 76,3 μV.

El decaimiento máximo que se le permite al S/H será: 76,3 μV/10 μs = 7,63 μV/μs.

La razón de atenuación de la entrada en la salida mínima será: $A_{Fmin} = 6(16 + 1) = 102$ dB para una frecuencia de hasta 1000 Hz.

22.6.3 LA TRANSICIÓN DEL MODO MUESTREO AL MODO RETENCIÓN

En el interruptor analógico S se producen una serie de fenómenos en su tránsito de abierto a cerrado que serán estudiados más adelante y, producto de ello, ocurre un transitorio previo al decaimiento de la salida, ilustrado en la figura 22.14.

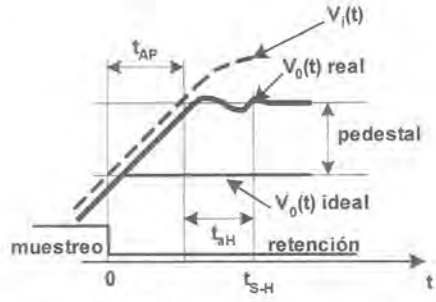


Figura 22.14. Transición del modo muestreo al modo retención.

Si en t = 0 la señal de control de modo pasa de muestreo a retención, el interruptor S tiene un tiempo de conmutación que provoca que la salida aún "siga" a la entrada durante un tiempo t_{AP} denominado **retardo de apertura** (*aperture delay*). Además, se produce una variación del valor de la señal capturada causada por la transferencia de carga hacia el condensador por el interruptor al abrirse que provoca una tensión de error llamada **paso o pedestal** (*sample-to-hold offset*) y un **transitorio de establecimiento** (*hold mode settling time*) (t_{aH} en la figura 22.14) de un valor final o zona plana que es el error de pedestal. Este error, mostrado de manera exagerada en la figura 22.14, debe ser menor de ½LSB del convertidor A/D que sigue al circuito S/H. Después de este transitorio comienza el decaimiento ilustrado en la figura 22.13.

De esta forma, la transición del modo muestreo al modo retención tendrá una duración $t_{S\text{-}H}$, dependiente fundamentalmente de la rapidez de conmutación del interruptor S, muy ligado a su tecnología de fabricación, y de los factores apuntados anteriormente.

Si se requiere una temporización exacta para disponer de la salida en el modo retención, será necesario considerar estos parámetros y, además, el retardo digital del circuito de control de modo.

EJERCICIO 22.4

Un circuito de muestreo y retención posee los siguientes parámetros:

Tiempo de apertura = 80 ns

Pedestal (step offset) = ±8 mV (independiente de la entrada)

Transitorio de conmutación:

– Amplitud = ±7,5 mV

– Tiempo de establecimiento hasta ±0,003% = 0,4 μs

– Tiempo de establecimiento hasta ±0,00076% = 1 μs

Decaimiento = 0,02 μV/μs.

Si su salida se entrega a un convertidor A/D de 8 bits, cuyo margen de entrada es de 0 ÷ 10 V y su tiempo de conversión de 200 μs, ¿con qué retraso respecto a la señal de control de modo del S/H se necesita comenzar la conversión en el convertidor A/D para evitar errores, considerando un retardo digital máximo de 50 ns? ¿Qué error producirá en la medida el pedestal del S/H? ¿Qué error producirá en la medida el decaimiento del S/H?

SOLUCIÓN:

El retraso con que debe comenzar la conversión respecto al instante de la transición del modo S al modo H de la señal de control de modo es el tiempo de apertura total.

Retraso = retardo digital + retardo de apertura + tiempo de establecimiento

En función de la resolución normalizada del convertidor A/D expresada en % (0,39% es la correspondiente a 8 bits) se puede escoger la duración del tiempo de establecimiento. En este caso se puede seleccionar 0,4 μs.

El retraso entonces será: 50 ns +400 ns + 80 ns = 0,53 μs.

La resolución aproximada del convertidor A/D será: 1 LSB = 10 V/2^8 = 39,06 mV y ½LSB = 19,53 mV. Como que el pedestal (8 mV) es menor que ½ LSB, no existirá error en la medida por este concepto.

La variación de la tensión en la salida en el modo H que provocará el decaimiento del circuito S/H durante el tiempo de conversión del convertidor A/D será:

$$\Delta V_0 = (0{,}02\ \mu V/\mu s)(200\ \mu s) = 4\ \mu V.$$

Como $\Delta V_0 \ll$ ½ LSB, el decaimiento no producirá error alguno en la medida.

Otro fenómeno importante es la **incertidumbre del tiempo de apertura** (*aperture jitter*). El tiempo de apertura t_{AP} no es constante, tiene cierta variación debido al ruido en la señal de control de modo, así como al desigual tiempo de conmutación de un mismo interruptor en diferentes momentos. Esta incertidumbre debe tenerse en cuenta pues puede provocar errores en el convertidor A/D que serán analizados en el capítulo 23.

22.6.4 CIRCUITOS PRÁCTICOS DE MUESTREO Y RETENCIÓN

Para evitar los efectos que sobre los parámetros característicos del sistema S/H tienen la carga y la resistencia de la fuente de señal, las arquitecturas básicas de los circuitos reales incorporan dos amplificadores operacionales como seguidores de tensión (*buffers*) de ganancia generalmente unitaria de manera tal que el conjunto interruptor electrónico-condensador no quede "aislado" de la entrada y de la salida. La figura 22.15 presenta tres de las arquitecturas más generalizadas en los circuitos prácticos de muestreo y retención [22.3]

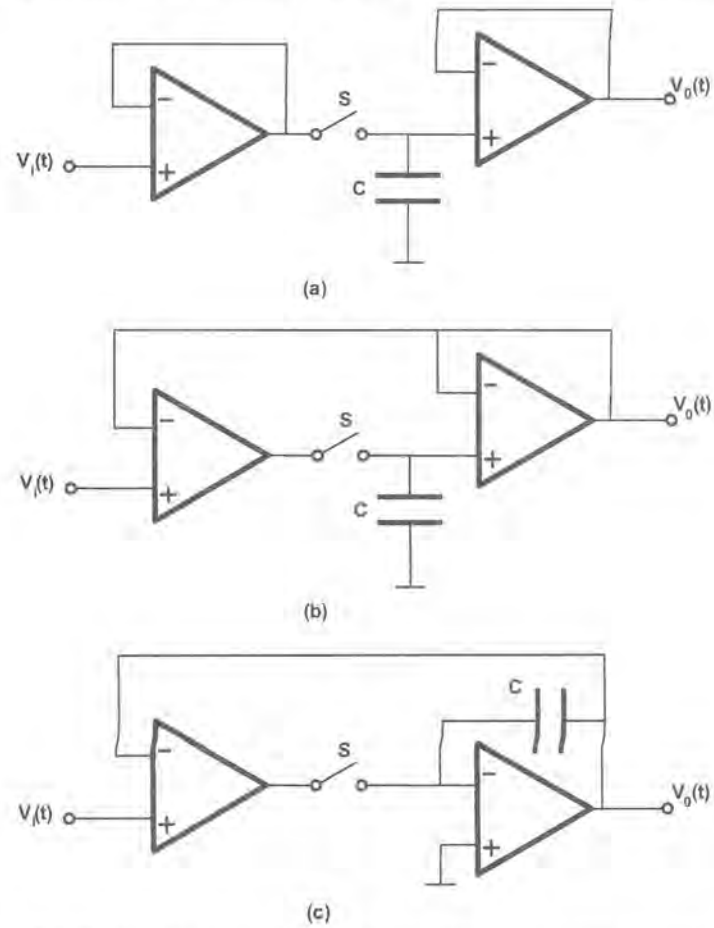


Figura 22.15. Arquitecturas básicas de circuitos de muestreo y retención.

El circuito de la figura 22.15a está en bucle abierto; los amplificadores son seguidores de tensión y no hay realimentación por lo que el tiempo de adquisición es muy pequeño. En consecuencia su exactitud no es buena ya que la ausencia de realimentación de la salida evita la reducción de las tensiones de *offset* de ambos amplificadores.

El circuito de la figura 22.15b está en bucle cerrado por la existencia de realimentación; la precisión es muy buena en detrimento del tiempo de adquisición debido a que, cuando el circuito se encuentra en el modo H, el primer amplificador se encuentra saturado y, al pasar al modo S, la salida de saturación ralentiza la operación de muestreo. Esta limitación puede ser salvable en alguna medida si, por ejemplo, se colocan de la salida del primer amplificador a su entrada inversora dos diodos en antiparalelo. La figura 22.16 presenta el diagrama simplificado del LF398, de tecnología BiFET y condensador externo que utiliza esta técnica.

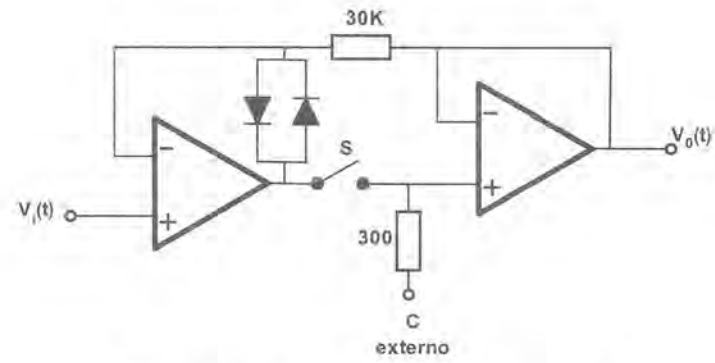


Figura 22.16. Diagrama simplificado del LF398.

En ambos circuitos la transferencia de carga desde el interruptor al abrirse y con ello la magnitud del pedestal dependen de la señal de entrada, ya que el condensador está conectado a esta a través del amplificador.

El circuito mostrado en la figura 22.15c reduce este problema al estar conectado el condensador a una masa virtual en lugar de a la entrada de señal con lo que no se evita la transferencia de carga que origina el pedestal, aunque es constante. No obstante, la salida es invertida respecto a la entrada. Los problemas relacionados con la saturación del primer amplificador son resueltos como en el circuito anterior, pero con el concurso de resistencias y un interruptor S_2 asociado al circuito de control de modo igual que S_1, como lo muestra la figura 22.17, donde los interruptores exhiben la posición correspondiente al modo H.

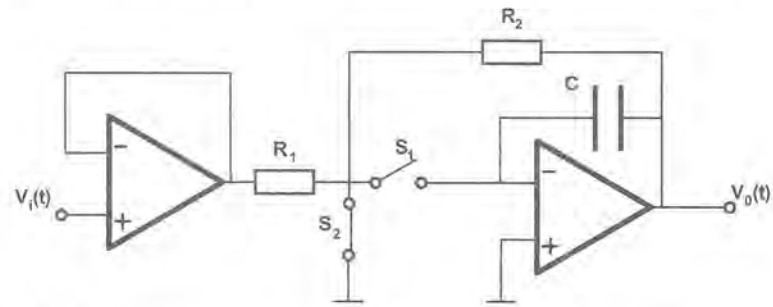


Figura 22.17. Configuración integradora sin saturación del amplificador de entrada.

Como los interruptores están conectados a masa, no existe una tensión de modo común en ellos, reduciendo la magnitud del pedestal. Esta arquitectura es muy utilizada en los convertidores analógico-digitales que incorporan el circuito S/H como parte del circuito integrado.

Para los circuitos S/H prácticos existen especificaciones vinculadas con la presencia de los amplificadores operacionales, tales como la tensión de *offset*, la ganancia, el error de ganancia y el ancho de banda (de interés sólo por el comportamiento del seguimiento salida-entrada durante el modo S), además de los que dependen específicamente del interruptor analógico y del condensador. La tabla 22.1 presenta una comparación entre los valores de los parámetros fundamentales de tres circuitos integrados de diferentes arquitecturas fabricados por Analog Devices y National Semiconductors [22.4]. Puede apreciarse que la arquitectura con integrador en la salida es la que presenta las mejores características estáticas y dinámicas.

PARÁMETRO	AD1154	LF398	LH4860
Arquitectura	Bucle abierto, amplificador	Bucle cerrado, amplificador	Bucle cerrado, salida a integrador
Tensión de desviación de entrada	±20 mV[1]	±2 mV	±0,5 mV
Error de ganancia	±0,01%	±0,004%	±0,005%
Ancho de banda en pequeña señal	1 MHz		16 MHz
Tiempo de apertura	80 ns	200 ns	6 ns
Incertidumbre de apertura	150 ps		35 ps[rms]
Tiempo de adquisición a 0,1%[2]	2,2 μs	4 μs	100 ns
Tiempo de establecimiento del modo H (0,01%)	200 ns	1 μs	60 ns
Decaimiento	±0,2 μV/μs		±0,5 μV/μs
Corriente de fugas del condensador		30 pA	5 pA
Razón de atenuación de la entrada en la salida, A_F a 1 kHz	106 dB[3]	90 dB	

[1] Ajustable a ±3 mV. [2]ΔV_0 = 10 V y C = 1000 pF en el LF398. [3]Medida a f = 10 kHz.

Tabla 22.1. Especificaciones fundamentales de algunos circuitos S/H integrados.

22.7 Multiplexación de señales analógicas

La multiplexación de señales analógicas es una función que se realiza en sistemas de adquisición de múltiples señales o canales analógicos que utilizan de forma común alguno de los componentes del canal, generalmente el S/H y/o el convertidor A/D. Consiste en conectar varias señales de entrada a una salida común, de manera tal que en la salida sólo esté presente **la entrada seleccionada** por el sistema que realiza esta función y que recibe el nombre de **multiplexor**, representado en la figura 22.18.

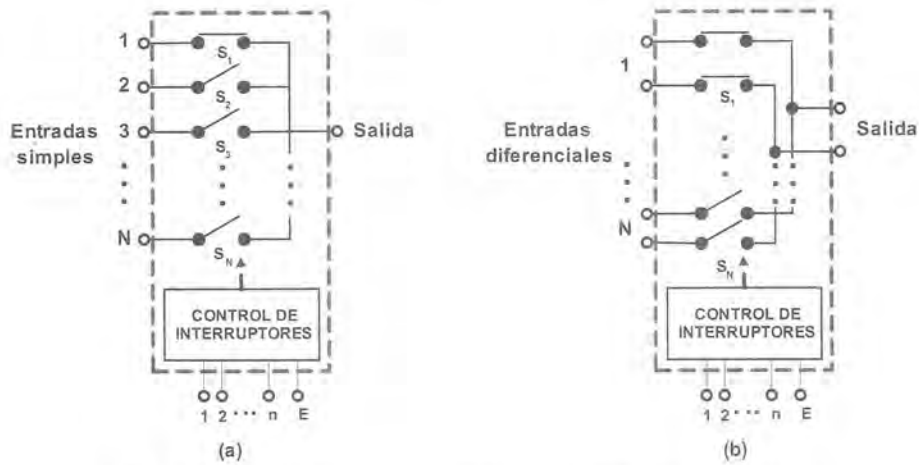


Figura 22.18. Diagrama general de un multiplexor de señales.

El circuito de control digital es el encargado de cerrar el interruptor que corresponda a la palabra binaria de n bits impuesta en sus entradas digitales de selección de canal (1,2,...n) cuando se activa su entrada digital de habilitación o "*enable*" (E). Idealmente, la entrada seleccionada debe estar presente en la salida inmediatamente después de ser activada la habilitación E.

El multiplexor representado en la figura 22.18a posee N canales y es utilizado para multiplexar N señales con una referencia común. Si las señales no tienen una referencia común, es decir, son diferenciales, la arquitectura utilizada es la superposición de dos multiplexores simples como el representado en la figura 22.18b, por lo que cada entrada es un par de terminales que son conectados correspondientemente al par de terminales de salida cuando el par de interruptores que accionan a la vez se cierran al ser seleccionados por un mismo código. Para aplicaciones en las que un conjunto de señales diferenciales pueden tener entre sí tensiones de modo común muy elevadas, los multiplexores diferenciales no se utilizan debido a que pueden ser destruidos por altas tensiones.

El "corazón" de un multiplexor es el interruptor analógico que comunica cada entrada con la salida común y éste, idealmente, debe presentar cero resistencia cuando está cerrado, infinita cuando está abierto y una transición entre estos estados muy rápida. En el Anexo 4 que se encuentra en el CD que se acompaña se ofrece una breve descripción del circuito equivalente de un interruptor analógico y los parámetros fundamentales asociados con su trabajo en los modos cerrado y abierto.

Los multiplexores analógicos prácticos consisten en un conjunto de interruptores analógicos CMOS fabricados como un circuito integrado que incorporan un circuito de control de los interruptores con o sin almacenamiento del código de selección y que eventualmente pueden incorporar algún *buffer* de salida, aunque este ralentiza los tiempos de conmutación, fundamentalmente el paso de abierto a cerrado. Estos circuitos están caracterizados por una serie de parámetros en gran medida determinados por el comportamiento estático y dinámico de los interruptores analógicos (véase Anexo 4).

Estos parámetros son:

- **Margen de señal analógica (*analog signal range*).** Expresa el margen de valores de la señal de entrada que el multiplexor es capaz de admitir. En los fabricados con tecnología LC^2MOS coincide con las fuentes $+V_{DD}$ y $-V_{SS}$ en funcionamiento bipolar.
- **Resistencia de conexión (R_{ON}).** Expresa el valor de la resistencia entre la entrada y la salida del interruptor seleccionado cuando está cerrado. Algunos fabricantes especifican su margen de valores correspondientes al margen de temperatura de trabajo u ofrecen gráficamente su dependencia con ella. Los valores típicos se encuentran entre 40 y 100 Ω.
- **Corrientes de fuga (*leakage currents*).** Expresan las magnitudes de las corrientes de fuga de los transistores MOS. Se encuentran en el orden de ±1 a ±20 nA en función de la temperatura de trabajo.
- **Tiempo de conmutación (*switching time*)**[10]. Expresa el tiempo necesario para que el interruptor seleccionado pase del estado de cierre al de apertura y viceversa con un error dado de aproximación al valor final (generalmente con una adquisición de un 90% del valor final). Sus valores pueden estar entre 50 y 800 ns.
- **Retardo de apertura antes de cierre (*break-before-make delay*;T_{OPEN}).** Es el retardo existente en el multiplexor entre el instante en que un interruptor ya está abierto y el instante en que otro ya se cierra, cuando ambos son accionados a la vez. Este atraso prevé que no exista la condición de ambos interruptores cerrados con sus drásticas consecuencias. Los valores típicos suelen estar entre 10 y 15 ns.

[10] Muchos fabricantes especifican además el tiempo de cierre (t_{ON}) y el de apertura (t_{OFF}), donde siempre el segundo es más pequeño que el primero.

- **Capacidades de entrada y salida C_{SOFF}, C_{DOFF}, C_{DON} y C_{SON}.** Son las capacidades que existen desde los terminales de entrada y de salida hasta la masa del circuito. Las mayores son las medidas en el estado cerrado y son del orden de decenas de pF.

- **Capacidad de entrada-salida C_{DS}.** Es la capacidad equivalente que existe entre la entrada y la salida con el interruptor abierto. Suele ser de pocos pF.

- **Aislamiento en el modo abierto (*OFF Isolation*).** Se refiere a la atenuación que recibe la señal aplicada en la entrada en su paso a través del interruptor abierto. Se expresa en decibelios, y es una función de la frecuencia de la señal de entrada y de la carga. Los valores típicos están alrededor de 70 dB a frecuencias del orden de cientos de kilohertzios.

- **Diafonía (*channel-to-channel crosstalk*).** Expresa la atenuación que ofrece el multiplexor a la señal de entrada aplicada a un canal seleccionado hasta los interruptores vecinos. Se expresa en decibelios y depende de la carga y de la frecuencia de la señal de entrada. Sus valores típicos son del orden de los del aislamiento en el modo OFF.

También son especificados los niveles mínimos y máximos de las señales digitales con niveles de tensión correspondientes al "1" y al "0" lógicos que el circuito de control de canales recibe, así como las duraciones mínimas de los pulsos para las señales de habilitación, las corrientes y capacidades de estas entradas lógicas y las tablas de asociación de canales-código de selección que, por lo general, utilizan el código binario directo.

En el CD que se acompaña se encuentran las hojas de características de los multiplexores de Analog Devices ADG406/07/26, de 8/16 canales y ADG428/29, de 4/8 canales con almacenamiento del código de selección, ambos de tecnología LC^2MOS.

22.7.1 LA MULTIPLEXACIÓN Y LA FRECUENCIA DE MUESTREO

La utilización de multiplexores en los sistemas de adquisición de datos está muy difundida para lograr el tratamiento eficiente de muchas señales analógicas de reducido espectro frecuencial por parte de sistemas de alto coste y amplias posibilidades como los amplificadores de instrumentación de ganancia programable, circuitos de muestreo y retención de alta calidad y gran rapidez y/o convertidores analógico-digitales muy rápidos.

En la figura 22.8 presentamos la estructura general de un canal simple de adquisición de una señal analógica en el que la señal está presente en la entrada del circuito de muestreo y retención, y la salida de este está conectada a la entrada del convertidor A/D **que necesita que la señal de entrada sea constante durante su tiempo de conversión**.

El tiempo mínimo en que se puede realizar la adquisición del valor de la señal analógica estará determinado por las siguientes duraciones:

$$T_{min} = (\text{tiempo de adquisición} + \text{tiempo de establecimiento} + \text{retardo de apertura})_{S/H} + \\ + \text{tiempo de conversión A/D}$$

En consecuencia, la frecuencia de muestreo será menor o igual al inverso de este tiempo, sin considerar el tiempo que el procesador emplee en hacer la lectura del dato.

Al utilizar un multiplexor analógico para el muestreo secuencial de un conjunto de señales, el precio que se paga es una reducción de la frecuencia de muestreo máxima que estará en dependencia de la arquitectura de adquisición que se utilice.

La figura 22.19 presenta un diagrama general de un sistema de **adquisición secuencial** con amplificador de instrumentación de ganancia programable para 8 canales.

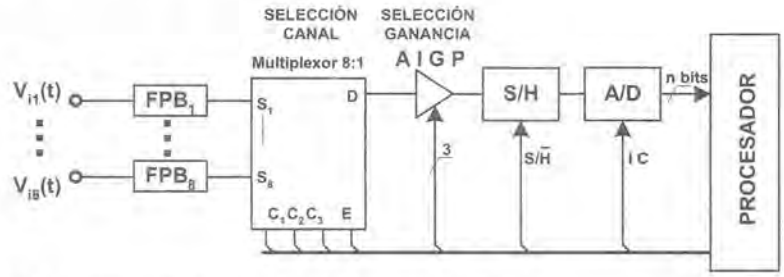


Figura 22.19. Sistema de adquisición secuencial de 8 canales con multiplexor analógico.

El procesador digital envía simultáneamente el código de selección del canal y de la ganancia correspondiente (si es posible se utilizan las mismas señales) y ordena al S/H que inicie su modo muestreo haciendo S/H = 1. Este solapamiento resulta posible y necesario para minimizar el tiempo de adquisición de la señal en el S/H. Transcurrido este "tiempo solapado" se ordena al S/H pasar al modo retención, haciendo S/H = 0. Transcurrido el retardo de apertura y el tiempo de establecimiento en el modo retención, se ordena al convertidor A/D iniciar la conversión haciendo IC = 1. Transcurrido el tiempo de conversión, el valor de la señal en un formato digital aparece estable en las salidas del convertidor A/D.

Sin apreciar el tiempo que el procesador emplea en dar entrada al dato y si llamamos:

- t_{MUX} al tiempo de conmutación del multiplexor analógico,
- t_{SAMPLE} al tiempo de adquisición + establecimiento del modo S del S/H,
- t_{HOLD} al retardo de apertura + tiempo de establecimiento del modo H del S/H,
- t_{AIGP} al tiempo de establecimiento del amplificador de ganancia programable y
- t_{CONV} al tiempo de conversión del convertidor A/D,

el tiempo mínimo necesario para realizar la adquisición de la señal de un canal será:

$$T_{min} \geq \sqrt{t_{MUX}^{\ 2} + t_{SAMPLE}^{\ 2} + t_{AIGP}^{\ 2}} + t_{HOLD} + t_{CONV} \qquad (22.10)$$

Este proceso se repite N veces, siendo N el número de canales del multiplexor, de manera tal que el periodo de muestreo mínimo para cada señal será:

$$T_{minCANAL} \geq T_{min} \cdot N \qquad (22.11)$$

Si $t_{MUX} + t_{AIGP}$ es mucho menor que t_{CONV}, una vez iniciada la conversión se puede enviar el código correspondiente al canal y ganancia siguientes correspondientes al multiplexor y al amplificador respectivamente, quedando de esta forma (22.10) reducida a:

$$T_{min} \geq t_{SAMPLE} + t_{HOLD} + t_{CONV} \qquad (22.12)$$

En la arquitectura mostrada, en el caso de no existir un amplificador de instrumentación debido a que la adecuación de los márgenes estáticos de las señales al margen de entrada del convertidor A/D se ha realizado en cada canal antes del filtrado, entonces (22.10) queda reducida a (22.12) debido a que $t_{MUX} << t_{SAMPLE}$.

Todos los tiempos de las expresiones anteriores son los correspondientes a la resolución con la que se desea hacer la adquisición y, además, hay que considerar que en la entrada del amplificador pueden estar presentes de forma consecutiva tensiones que precisan ganancias muy diferentes por lo que los tiempos de establecimiento son distintos y se debe seleccionar el mayor de todos.

EJERCICIO 22.5

Un sistema de adquisición de datos secuencial de 8 canales analógicos utiliza los siguientes dispositivos: convertidor A/D de 12 bits, (0 ÷10 V) con t_{CONV} = 30 μs, circuito de muestreo y retención con t_{SAMPLE} = 3,5 μs y t_{HOLD} = 1 μs; amplificador de ganancia programable por software con t_{AIGP} = 15 μs y multiplexor con t_{MUX} = 350 ns. Calcule la máxima frecuencia de muestreo para cada canal.

SOLUCIÓN:

La duración mínima del ciclo de muestreo de un canal con estos dispositivos en el sistema estará dada por:

$$T_{min} = \sqrt{t_{MUX}^2 + t_{SAMPLE}^2 + t_{AIGP}^2} + t_{HOLD} + t_{CONV} = \sqrt{0,35^2 + 3,5^2 + 15^2} + 1 + 30 = 46,41\ \mu s$$

y el tiempo de muestreo correspondiente a cada canal será:

$$T_{minCANAL} = 8T_{min} = 371,28\ \mu s$$

La frecuencia de muestreo de cada canal será:

$$f_s = 1/T_{minCANAL} = 1/371,28\ \mu s = 2,69\ kHz$$

Si se deseara reproducir las señales a partir de sus muestras y se considera que la frecuencia máxima del espectro no debe ser mayor que $f_s/10$ para no imponer grandes exigencias dinámicas a los filtros *anti-aliasing*, f_m quedaría limitada a:

$$f_{mmax} = 269\ Hz$$

Esta arquitectura tiene como ventaja la minimización de componentes del sistema, pero tiene el inconveniente de no disponer de valores de las señales medidas al mismo tiempo. Si esto último fuera una exigencia, entonces se utilizaría la arquitectura de **captura simultánea**, ilustrada en la figura 22.20.

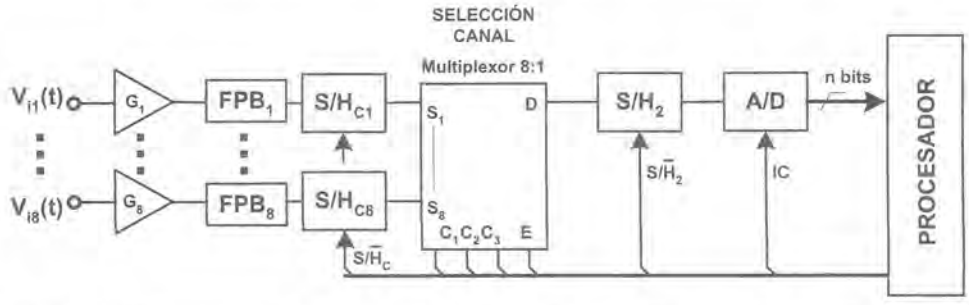


Figura 22.20. Sistema de adquisición simultánea de 8 canales con multiplexor analógico.

Se diferencia de la anterior en que, por lo general, en esta arquitectura no se utiliza un amplificador de ganancia programable; la adecuación de márgenes estáticos de las señales se realiza en cada canal. Además del S/H_2, asociado al convertidor A/D en el caso en que el convertidor utilizado lo necesite, se utilizan sistemas de muestreo y retención para cada canal (S/H_C) para lograr la captura simultánea de todas las señales.

El proceso de adquisición se inicia con los comandos muestreo ($S/H_C = S/H_2 = 1$) a todos los S/H, a la vez que se envía el código correspondiente al primer canal al multiplexor. Transcurrido $t_{SAMPLEC}$, se envía el comando retención ($S/H_C = 0$) a todos los S/H de los canales el cual es mantenido durante la adquisición

secuencial de todos los canales. Esta cabecera se añade al proceso anterior de adquisición secuencial sin incorporar el amplificador de instrumentación.

La expresión para el tiempo mínimo de adquisición de la señal de un canal será ahora:

$$T_{min} \geq \sqrt{t_{MUX}^{\ 2} + t_{SAMPLE_2}^{\ 2}} + t_{HOLD2} + t_{CONV} \tag{22.13}$$

y como siempre $t_{MUX} << t_{SAMPLE}$ y, si se cumple además que $t_{SAMPLE} >> t_{HOLD}$, entonces (22.13) quedaría aproximadamente como:

$$T_{min} \geq t_{SAMPLE2} + t_{CONV} \tag{22.14}$$

El periodo mínimo de muestreo de cada señal será:

$$T_{minCANAL} \geq (t_{SAMPLE} + t_{HOLD})_C + T_{min} \cdot N \tag{22.15}$$

La frecuencia de muestreo de cada canal será:

$$f_{SC} = 1/(t_{SAMPLEC} + t_{HOLDC} + T_{min} \cdot N) \tag{22.16}$$

Un aspecto muy importante a tener en cuenta en esta arquitectura es que el decaimiento que sufren las señales en los S/H de canales durante el modo retención debe ser menor que el peso de ½LSB para evitar errores en la conversión A/D.

EJERCICIO 22.6

Un sistema de adquisición de datos en arquitectura de captura simultánea utiliza los dispositivos del ejercicio 22.5, excepto el amplificador. Si el decaimiento en los S/H idénticos es de ±2 μV/μs, calcule la máxima frecuencia de muestreo para cada canal; ¿se produce algún error en la conversión debido al decaimiento en los S/H_C?

SOLUCIÓN:

Para estos dispositivos se cumple que $t_{MUX} = t_{SAMPLE}/10$, sin embargo, no se cumple que $t_{SAMPLE} >> t_{HOLD}$ ya que la resolución del convertidor exige un tiempo de establecimiento en retención importante.

El tiempo mínimo para la adquisición de un canal será:

$$T_{min} = t_{SAMPLE2} + t_{HOLD2} + t_{CONV} = 3{,}5 + 1 + 30 = 34{,}5\ \mu s$$

El periodo mínimo de muestreo de cada canal será:

$$T_{minCANAL} = (t_{SAMPLE} + t_{HOLD})_C + N \cdot T_{min} = 3{,}5 + 1 + 8(34{,}5) = 280{,}5\ \mu s$$

La frecuencia de muestreo de cada canal será:

$$f_{SC} = 1/T_{minCANAL} = 1/280{,}5\ \mu s = 3{,}565\ kHz$$

El canal más crítico respecto al decaimiento de su S/H es el último que se seleccione en el ciclo de muestreo y, en el peor de los casos, aproximadamente ese circuito se mantendrá en el modo retención durante su periodo de muestreo, menos el tiempo de adquisición por parte de su S/H.

Así las cosas, $\Delta T_{HOLDmáx} = 280{,}5 - 3{,}5 = 277\ \mu s$.

Durante este tiempo, la caída de tensión en el condensador del S/H_C será:

$$\Delta V_{máx} = (2\ \mu V/\mu s)(277\ \mu s) = 0{,}554\ mV$$

Y el peso de ½ LSB para el convertidor aproximadamente es: ½ LSB = 10 V/2^{n-1}= $10/2^{13}$ = 1,2 mV.

Como $\Delta V_{máx}$ < ½ LSB, para esta frecuencia de muestreo de cada canal el decaimiento no produce error en la conversión.

Resumen

- El resultado del muestreo de una variable analógica es una variable muestreada, formada por los valores que tiene la variable analógica cada cierto intervalo regular de tiempo T, denominado periodo de muestreo. En la medida que el periodo de muestreo sea menor, mejor será la relación de identidad existente entre la variable muestreada y la analógica.
- Si se desea reconstruir una señal analógica de ancho de banda finito a partir de sus muestras sin pérdida de información, la frecuencia de muestreo tiene que ser, al menos, el doble de la máxima frecuencia del espectro frecuencial de la señal analógica original. Si el muestreo se realiza con una frecuencia menor que la de Nyquist ($f_s = 2f_m$), se producirá un solapamiento entre los espectros de la señal y las componentes de la modulación de amplitud que tiene lugar por el proceso de muestreo, dando como resultado la consiguiente pérdida de información.
- En los sistemas de muestreo reales el *aliasing* es evitable utilizando filtros *anti-aliasing* que garanticen que la amplitud de las componentes espectrales solapadas sea menor que la resolución del sistema convertidor A/D que sucede al muestreador. Las exigencias sobre estos filtros son mayores cuanto más cerca estén f_s y f_m y, como criterio práctico, se aconseja que la frecuencia de muestreo nunca sea menor que $5f_m$.
- La reconstrucción de señales a partir de sus muestras se realiza con los convertidores digital-analógicos, seguidos de filtros para eliminar las componentes espectrales no deseadas de la señal muestreada.
- El circuito de muestreo y retención (S/H) realiza la función de muestreador de la señal analógica y, a la vez, retiene a su salida el valor de la muestra durante un tiempo dado. Tanto en el seguimiento de la señal analógica de entrada durante el modo muestreo como en la retención del valor adquirido durante el modo retención, existen fuentes de error que deben ser cuidadosamente evaluadas en función de la resolución del convertidor A/D que le sucede.
- Un multiplexor analógico es un circuito electrónico formado por un conjunto de interruptores analógicos que conectan la entrada del interruptor seleccionado a la salida común, mientras que el resto permanece en el modo de no conducción.
- La aplicación de los multiplexores analógicos en los sistemas de adquisición de datos con muestreo secuencial o simultáneo optimizan el uso de los circuitos de conversión analógica-digital, pero reducen la frecuencia de muestreo de cada canal en proporción directa al número de canales.

Problemas propuestos

22.1 Una señal de tensión triangular periódica de frecuencia 500 Hz y amplitud ± 5 V se desea muestrear de manera tal que la tensión entre dos muestras consecutivas no exceda a 1 mV, que es la resolución del convertidor A/D del sistema. ¿Cuál es la mínima frecuencia de muestreo que satisface estas

condiciones? Si se muestrea con una frecuencia de 2000 Hz, ¿se podría reconstruir la señal a partir de sus muestras? Fundamente su respuesta.

22.2 Se desea limitar el ancho de banda de una señal analógica de margen estático 0 ÷5 V y gran espectro frecuencial hasta la frecuencia de 1000 Hz. El convertidor A/D es de 12 bits y su margen estático de entrada es 0 ÷5 V. Especifique los parámetros del filtro *anti-aliasing* necesario si: a) $f_s = 5$ kHz, b) $f_s = 10$ kHz.

22.3 En el problema anterior, considere que el convertidor A/D es de 16 bits.

22.4 La salida de un circuito S/H cuyas características se detallan en el Ejercicio 22.4 se entrega a un convertidor A/D de 12 bits y margen de entrada 0 ÷5 V. Si el retardo digital es de 50 ns respecto a la señal de control de modo, ¿con qué retardo se debe iniciar la conversión en el convertidor A/D respecto a esta señal? ¿Qué error produce en la medida el pedestal del S/H?

22.5 Si el circuito S/H del problema anterior tiene un decaimiento de ± 10 μV/μs, ¿cuál es el máximo tiempo de conversión del convertidor A/D que no produce error en la conversión? Calcule la frecuencia de muestreo para conversiones consecutivas con este tiempo de conversión.

22.6 Un multiplexor analógico de 8 canales simples ADG428 con fuentes de alimentación de ± 15 V (sus hojas de características se encuentran en el CD adjunto) se utiliza en un sistema de adquisición de datos. El margen estático de las señales conectadas a sus entradas es 0,5 V. Si la carga equivalente conectada a su salida puede considerarse como una resistencia de 1 kΩ en paralelo con un condensador de 35 pF, calcule el tiempo de transición abierto-cerrado necesario para tener un error de aproximación en la salida de 0,001%. Sugerencia: Utilice el Anexo 4 en el CD adjunto. Si el sistema utiliza los componentes del problema 22.4 con el tiempo de conversión del CA/D calculado en el problema 22.5, su arquitectura es de captura consecutiva, calcule la frecuencia de muestreo de cada canal.

22.7 Repita el problema 22.5 considerando que el procesador emplea 8 μs en leer el dato del convertidor A/D.

22.8 Un sistema de adquisición de datos de captura simultánea utiliza las componentes del Ejercicio 22.6. a) Calcule la máxima frecuencia de muestreo de cada canal considerando que el procesador emplea 8 μs en leer el dato del convertidor A/D. Compare los resultados obtenidos con el problema anterior. Evalúe el error producido por el decaimiento de los S/H de canal.

23

Conversión entre variables analógicas y digitales

23.1 Cuantificación
23.2 Codificación
23.3 Conversión Digital-Analógica
23.4 Conversión Analógica-Digital

Las variables analógicas del proceso (después de convertidas en señales eléctricas) requieren una transformación a un **formato digital** para ser tratadas por un procesador digital. Esto significa que cada valor de la señal tiene que estar representado por un código diferente y para ello, el sistema de medida necesita una interfaz capaz de realizar esa función: el convertidor Analógico-Digital (CA/D).

El resultado de la medición elaborado en el procesador digital existe en forma de un código. Para utilizar esta información como **señal eléctrica** se necesita de una interfaz capaz de asociar a cada código de su entrada una tensión o corriente a su salida: el convertidor Digital-Analógico (CD/A).

23.1 Cuantificación

Cuantificar una señal eléctrica continua consiste en **representar** sus valores en forma de una serie finita de N niveles de tensión, corriente o estados diferentes. El sistema que realiza esta función recibe el nombre de **cuantificador**. La figura 23.1 presenta la característica transferencial estática ideal de un cuantificador de N niveles o estados, el error de cuantificación y su densidad de probabilidad.

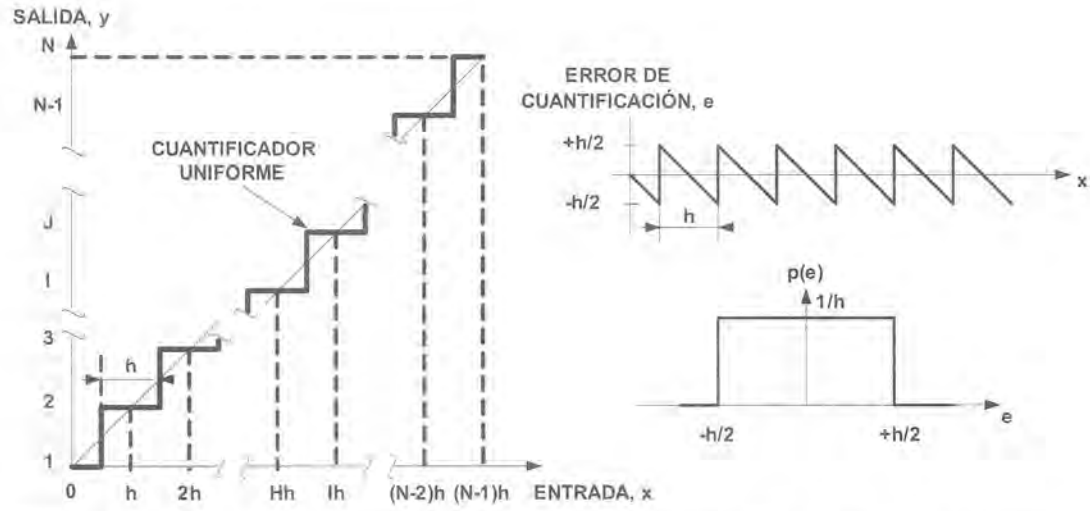


Figura 23.1. Función de transferencia de un cuantificador uniforme y su error de cuantificación.

El margen de variación de la señal de entrada **h** para el cual la salida no cambia de valor o de estado se denomina **intervalo de cuantificación** y es equivalente a la **resolución** del cuantificador. Para un cuantificador de alcance unitario en el que el número de estados o niveles diferentes de salida sea N, la resolución se expresará como:

$$\text{Resolución} = 1/(N-1)$$

Pero si N es muy grande[1] en comparación con 1, entonces aproximadamente se suele expresar como:

$$\text{Resolución} \approx 1/N \tag{23.1}$$

Si la resolución es la misma para todo el margen estático de entrada del cuantificador de fondo de escala FS, entonces estamos en presencia de un **cuantificador uniforme** para el cual el intervalo de cuantificación está fijado por:

$$h = \frac{FS}{N} \tag{23.2}$$

Inherente a la función del cuantificador real existe el **error de cuantificación** cuya amplitud o valor máximo se corresponde con la mitad del valor del intervalo de cuantificación (h/2). En la cuantificación uniforme este error varía con la señal de entrada en forma de un diente de sierra de periodo h (figura 23.1) por lo que, si N es grande y la señal de entrada se encuentra dentro del margen de medida del cuantificador, la función densidad de probabilidad p(e) tiene forma rectangular de amplitud 1/h en todo el intervalo $-h/2 \leq e \leq +h/2$.

Si expresáramos N en forma de una potencia de base 2, es decir, $N = 2^n$ (donde n es el número de bits del código binario directo con que se representa a N), podemos conocer la **resolución** del cuantificador uniforme **en función de n**. Así las cosas, (23.1) y (23.2) pueden expresarse como:

$$\text{Resolución} = \frac{1}{2^n} \tag{23.3}$$

[1] Físicamente la resolución es $1/(N-1)$. Pero si $N = 2^n$, la diferencia entre $1/2^n$ y $1/(2^n-1)$ es de un 3% para $N = 32$ ($n = 5$) y de un 0,39% cuando $N = 256$ ($n = 8$). Por ello para los cálculos se suele hacer esta aproximación.

$$h = \frac{FS}{2^n} \tag{23.4}$$

Cuando la salida **y** es expresada en términos de estados y estos son representados por N códigos digitales diferentes, la diferencia entre los códigos correspondientes a dos estados adyacentes (*I*, *J*) es la unidad. Esta diferencia, expresada en términos de la señal de entrada **x**, que generalmente es una tensión, se denomina **peso del LSB** (***Least Significant Bit***) y es equivalente a h.

En la tabla 23.1 se presenta el valor de la resolución aproximada en función del número de estados. Obsérvese que en la medida en que n crece, la resolución es menor y, por ende, menor el error de cuantificación.

n, bits	Número de estados, 2^n	Resolución = $1/2^n$	
		%	ppm
2	4	25	250.000
4	16	6,25	62.500
8	256	0,390625	3.906,25
12	4096	0,024414	244,14
16	65536	0,001526	15,26
24	16777216	0,000006	0,06

Tabla 23.1. Resolución aproximada del cuantificador uniforme.

El error de cuantificación máximo puede ser expresado como:

$$e_{max} = \pm\frac{1}{2}\cdot\frac{1}{2^n} \tag{23.5}$$

EJERCICIO 23.1

a) ¿Qué número mínimo de niveles garantiza un error de cuantificación máximo de ±1% en un cuantificador uniforme?

Si los niveles de salida se expresan mediante códigos correspondientes al binario directo,

b) ¿De cuántos bits necesitará ser el cuantificador?

c) ¿Qué resolución tiene?

d) ¿Cuál es el error máximo de cuantificación?

SOLUCIÓN:

a) De la ecuación (23.5) se tiene:

$\pm 0{,}01 \geq \pm 1/2N$ y de la condición "igual a" se obtiene **N = 50**.

Si se expresaran estos niveles mediante códigos en binario directo, $2^n > 50$, donde n es entero.

b) El número de bits es el menor número entero que satisface la condición n = 6 ($2^6 = 64$). La salida del cuantificador será representada por códigos de 6 bits. El cuantificador será de 6 bits.

c) y d) Los valores exactos de la resolución y el error máximo de cuantificación para un cuantificador de 6 bits serán:

$$\%\ \text{resolución}_e = (1/(2^6\text{-}1))\cdot 100\% = 1{,}5873\%;\ e_{máx} = \pm(1/2\cdot 63)\cdot 100\% = \pm 0{,}7936\%.$$

Los valores aproximados de estos parámetros (expresiones 23.3 y 23.5) serán:

$$\%\ \text{resolución}_a = (1/2^6)\cdot 100\% = 1{,}5625\%;\ e_{máx} = \pm(1/2\cdot 64)\cdot 100\% = \pm 0{,}7943\%.$$

Obsérvese que la diferencia entre estos resultados es del orden de 1,5%.

En el cuantificador uniforme, el error de cuantificación puede interpretarse como si la señal de entrada tuviera un ruido aleatorio de función densidad de probabilidad rectangular y acotado en magnitud entre ±h/2. A este ruido le denominaremos **ruido de cuantificación** cuyo valor medio cuadrático viene dado por:

$$e^2 = \int_{-\infty}^{+\infty} e^2 \cdot p(e)\cdot de = \frac{1}{h}\int_{-\frac{h}{2}}^{+\frac{h}{2}} e^2 \cdot de = \frac{1}{h}\left[\frac{e^3}{3}\right]_{-h/2}^{h/2} = \frac{h^2}{12} \tag{23.6}$$

Obsérvese que la única forma de reducir este ruido es reduciendo el intervalo de cuantificación h, es decir, aumentando el número de estados N.

Si consideramos a la entrada del cuantificador una señal senoidal cuya amplitud coincida con el campo de medida S (de la figura 23.1, S = (N-1)h), su valor medio cuadrático será:

$$s^2 = \frac{1}{2\pi}\int_0^{2\pi}\left(2^n-1\right)^2\cdot h^2\cdot \text{sen}^2\theta\cdot d\theta = \left(2^n-1\right)^2\frac{h^2}{2} \tag{23.7}$$

y la relación señal/ruido para el cuantificador uniforme con entrada senoidal será:

$$S/N = \frac{\left(2^n-1\right)^2\frac{h^2}{2}}{\frac{h^2}{12}} = 6\cdot\left(2^n-1\right)^2 \tag{23.8}$$

La expresión (23.8) puede simplificarse si n es grande (basta con que n sea mayor que 6) y en forma aproximada resultar:

$$S/N = 6\cdot 2^{2n}$$

$$S/N = 10\log 6 + 20n\log 2 = 7{,}782 + (6{,}021)n\ \text{dB} \tag{23.9}$$

Utilizando este procedimiento puede calcularse la relación S/N máxima para diferentes formas de variar con el tiempo de la señal de entrada, considerando siempre que su valor máximo coincida con el valor de la entrada a fondo de escala, es decir, con $(2^n - 1)h$. Por ejemplo, para una entrada continua unipolar (muy propia de señales de procesos lentos, tales como la temperatura), el valor medio cuadrático de esta señal será $(2^n - 1)^2\cdot h^2$ y la relación S/N máxima será 10,792 + (6,021)n, en dB.

En el cuantificador uniforme la relación S/N es dependiente de la amplitud de la señal y, para amplitudes pequeñas, puede resultar muy poco favorable. Una solución a este problema es lograr que el cuantificador presente un intervalo de cuantificación h **no uniforme**, de manera tal que la magnitud del intervalo de cuantificación h sea una función directa de la amplitud de la señal. Con ello se consigue una relación S/N

aproximadamente constante, independiente de la señal [23.1]. Una realización con amplificadores logarítmicos y un cuantificador uniforme se presenta en [23.2].

23.2 Codificación

Codificar es representar biunívocamente mediante varios dígitos los diferentes estados o niveles finitos de un cuantificador. Por ejemplo, en el cuantificador de la figura 23.1, hemos representado los N estados o niveles de salida mediante un código **decimal**[2] sin incluir el cero.

La generalidad de los procesadores digitales interpretan ***códigos binarios*** *en los que las cifras utilizadas* sólo son dos, el cero (0) y el uno (1) y a la cifra se le denomina bit (***Binary Digit***).

Los códigos binarios que se utilizan para representar los estados correspondientes a señales unipolares reciben el nombre de **códigos binarios unipolares** y los que representan a señales bipolares **códigos binarios bipolares**.

23.2.1 CÓDIGOS BINARIOS UNIPOLARES

Los códigos binarios unipolares más comúnmente utilizados son el **binario directo** y el **decimal codificado en binario o BCD** *(Binary-Coded Decimal)*, aunque existen otros códigos como el Gray o binario reflejado de utilización muy específica en la solución de problemas de ambigüedad entre posiciones contiguas en codificadores de posición absolutos [23.3], al aprovechar la propiedad de este código de ser de distancia unidad, es decir, que entre dos códigos adyacentes sólo cambia un bit.

Binario directo o natural: en este código cada cifra (bit) tiene asociada una potencia de 2, en concordancia con su posición en la palabra digital. Para un código de n bits, la expresión general del código binario directo es la siguiente:

$$b_{n-1}b_{n-2}...b_i...b_1b_0 = b_{n-1}\cdot 2^{n-1} + b_{n-2}\cdot 2^{n-2} + ... + b_i\cdot 2^i \ldots + b_1\cdot 2^1 + b_0\cdot 2^0 \text{ donde } b_i \in [0,1]$$

En este código, al bit del extremo izquierdo o bit de mayor "peso" se le denomina bit más significativo o MSB (***Most Significant Bit***) y al del extremo derecho bit menos significativo o LSB (***Least Significant Bit***).

Por ejemplo, para representar 011101 en decimal, se tendría que:

$$011101 = 0\cdot 2^5 + 1\cdot 2^4 + 1\cdot 2^3 + 1\cdot 2^2 + 0\cdot 2^1 + 1\cdot 2^0 = 29$$

EJERCICIO 23.2

¿Cuál es el código binario directo que corresponde al 321?

SOLUCIÓN:

Si realizamos la división sucesiva del número decimal por 2, los restos que se obtengan serán los valores de los bits del código binario. El primer resto se corresponde con el bit de menor peso del código (coeficiente de 2^0 o LSB) y el último al de mayor peso (coeficiente de 2^{n-1} o MSB).

[2] El sistema numérico decimal o denario utiliza 10 como base o raíz, de modo que cualquier entero positivo puede ser representado como una serie de potencias de 10 o decenas: $d_{n-1}10^{n-1} + ... + d_i10^i + ... + d_110^1 + d_010^0$ donde $d_i \in [0,1,2...8,9]$.

Operación	321/2	160/2	80/2	40/2	20/2	10/2	5/2	2/2	1/2
Resto	1(LSB)	0	0	0	0	0	1	0	1(MSB)

El código será: 101000001.

EJERCICIO 23.3

¿Cuántos bits debe tener un código para poder representar 579 estados en un cuantificador?

SOLUCIÓN:

Conocemos que con un código binario directo de n bits se pueden representar N estados, donde $N = 2^n$. Esto puede escribirse como

$$n = \ln N/\ln 2. \text{ Si } N = 579, \text{ entonces } \ln 579/\ln 2 = 9{,}18$$

Por lo que el código binario directo que necesitamos tiene que ser de n = 10 bits ya que hay que redondear por exceso al entero inmediato.

Se utilizarían de este código las primeras 579 palabras binarias de las $2^{10} = 1024$ posibles.

Obsérvese que hemos nominado al primer estado como Estado 0 y al último como Estado 578.

0000000000→Estado 0
0000000001→Estado 1
0000000010→Estado 2
...
1001000000→Estado 577
1001000001→Estado 578

Decimal codificado en binario(BCD): en este código, cada dígito se codifica por separado en binario directo, por lo que hacen falta cuatro bits para representar los números del 0 al 9. Los bits se agrupan en cuartetas (*nibbles*) separadas por un espacio y cada *nibble* representa una cifra BCD. Cuando el peso de cada bit dentro del *nibble* es el mismo que en el código binario directo (8-4-2-1 para $b_3b_2b_1b_0$, respectivamente) se dice que el código es NBCD (BCD Natural). Aunque el NBCD es el más utilizado existen otros pesos para los bits de cada *nibble*, por ejemplo, el 2-4-2-1 y el 5-4-2-1.

Por ejemplo, si se quiere representar 347 en NBCD, sería:

$$347 = \underset{3}{\underline{0011}}\ \underset{4}{\underline{0100}}\ \underset{7}{\underline{0111}}$$

Estos códigos utilizan 10 combinaciones de las 16 posibles en cada *nibble* para representar las cifras decimales del 0 al 9 por lo que son poco eficientes. Sin embargo, se utilizan para trabajo con presentadores numéricos decimales y existen circuitos integrados que realizan operaciones aritméticas con estos códigos. La mayoría de los procesadores modernos permiten operaciones en NBCD.

23.2.2 CÓDIGOS BINARIOS BIPOLARES

Los códigos binarios bipolares más utilizados son el **binario con signo**, el **binario en complemento a uno**, el **binario en complemento a dos** y el **binario desplazado** o **binario con *offset***.

Binario con signo: en este código los valores se representan igual que en el binario natural, pero se le añade al código un bit a la izquierda de su bit más significativo. Este bit representa al **signo**, tomándose el **0** para el positivo y el **1** para el negativo.

Por ejemplo, si queremos representar 321 y –321 en binario con signo, tendríamos que del Ejercicio 23.2, la representación en binario natural de 321 es 101000001. En binario con signo, +321=**0**101000001; –321=**1**101000001.

Obsérvese que este es un código simétrico respecto al cero que tiene dos posibles representaciones. Se *suele utilizar sólo una de ellas, la correspondiente al 0+, es decir, con un cero delante del código binario* correspondiente al cero con n bits. No es un código muy utilizado debido a que dificulta mucho las operaciones aritméticas.

Binario en complemento a uno: en este código los valores positivos se representan como en binario con signo, pero los negativos no. Para estos se mantiene el bit de signo (1), y se realiza el complemento a uno del código, que es el código que resulta de cambiar todos los bits del código de origen por el complementario (el 0 por el 1 y viceversa). Si en el código binario directo se pueden adicionar tantos ceros a la izquierda como se desee, ahora no. *Para igualar el número de bits de un número negativo con el de otro positivo representado en un mayor número de bits, en el número negativo se adicionan "unos" (1) a la izquierda.* De igual forma que en el binario con signo, existen dos códigos para representar el cero: 00...00 y 11...11. La palabra digital de mayor peso positivo de este código es 011...11, mientras que la de mayor peso negativo será 100...00. Este código ofrece la ventaja de poder hacer restas directamente, aunque con la necesidad de realizar una corrección del resultado si es positivo.

Por ejemplo, supongamos que queremos representar 247 y –247 en binario con complemento a uno. La representación en binario directo de 247 (ver procedimiento del Ejercicio 23.2) es 11110111 y su complemento a 1 será: 00001000. Por lo que en binario con complemento a uno resultará:

+247→**0**11110111
-247→**1**00001000

EJERCICIO 23.4

Calcular la suma algebraica de 18 y –6 en complemento a uno.

SOLUCIÓN:

La representación en complemento a uno de 18 es **0**10010 (5 bits y el de signo) y la de –6 en igual número de bits es **1**11001. Obsérvese que se han añadido dos unos a la izquierda.

+18→**0**10010
-6→**1**11001

La suma binaria de todos los bits (incluidos los de signo) es: 001011 y hay un acarreo que es necesario sumar al resultado para obtener el valor correcto. Al sumar 1 a este resultado se obtiene: 001100. El bit del extremo izquierdo es cero con lo que el número en complemento a uno resultante en 5 bits es positivo y su representación se corresponde con +12.

Binario en complemento a dos: en este código los valores positivos se representan como en binario con complemento a uno, pero los negativos son el complemento a dos del código que representa al positivo. El complemento a dos de un número es su complemento a uno más uno. De esta forma, cuando se suman dos

números iguales de signo opuesto representados en complemento a dos el resultado es cero. Para encontrar fácilmente el código en binario directo que le corresponde a la representación en n bits de un número negativo –N en complemento a dos, puede utilizarse la siguiente regla:

$$[-N]_{COMP2} = \{2^n - [N]_{DEC}\}_{BIN}$$

Este código tiene la ventaja sobre el binario con complemento a uno de que las sumas algebraicas en complemento a dos se realizan directamente.

Supongamos que se desea representar 247 y –247 en binario con complemento a dos. La representación de 247 en binario directo es 011110111 y su complemento a dos será: $(2^9 - 247)_{BIN} = (265)_{BIN} = 100001001$. Por lo que en complemento a dos la representación de estos números será:

+247→**0**11110111
–247→**1**00001001

Obsérvese que el bit del extremo izquierdo representa el signo y que la suma de ambos es cero. Para representar el cero en 9 bits se usa el código **0**00000000.

EJERCICIO 23.5

Calcular la suma en complemento a dos de 18 y –6.

SOLUCIÓN:

La representación en complemento a dos de 18 es **0**10010, en 6 bits.

La representación de –6 en complemento a dos en 6 bits es **1**11010.

La suma algebraica será: 001100 y hubo acarreo que se desprecia si el resultado es positivo. El sexto bit es cero con lo que el resultado es positivo y los 6 bits de datos conforman el código que se corresponde con el 12.

Binario desplazado o binario con offset: este código es una versión del binario con complemento a dos en la que el bit de signo es opuesto. Ahora, para las cantidades positivas el bit de signo es 1, y 0 para las negativas. Al igual que en el binario con complemento a dos, la suma de dos números iguales de signos opuestos es cero y el cero se considera una cantidad positiva. La realización práctica en los convertidores de código a señal eléctrica (convertidores digital-analógicos) para señales bipolares utilizando este código resulta muy cómoda técnicamente, siendo la preferida por los fabricantes de estos sistemas en forma de circuito integrado.

Existen otros códigos como el octal y el hexadecimal a los que se hace una breve referencia en el Anexo 5 que se encuentra en el CD.

23.3 Conversión Digital-Analógica

La conversión digital-analógica es un procedimiento a través del cual un código de entrada φ es transformado en una señal de tensión o de corriente unipolar o bipolar de salida mediante una correspondencia entre 2^n

combinaciones binarias posibles de entrada y 2^n tensiones (o corrientes) discretas de salida, obtenidas a partir de una referencia de tensión o de corriente. En términos de tensión y de corriente, puede expresarse como:

$$V_o = \sum_{i=0}^{n-1} V_{pondi} \cdot dp_i \ , \text{donde } dp \in [0,1]$$

$$I_o = \sum_{i=0}^{n-1} I_{pondi} \cdot dp_i \ , \text{donde } dp \in [0,1] \qquad (23.10)$$

Este proceso consiste en hacer una ponderación de cada bit del código N y sumar el resultado de todas las ponderaciones multiplicadas por el valor del bit asociado a cada una. De esto se deriva que la salida (V_o o I_o) es única para cada código de entrada I. La salida de un convertidor digital-analógico de un número finito de códigos de entrada no es una señal analógica, sino discreta, digital.

23.3.1 CARACTERÍSTICAS DE LA CONVERSIÓN D/A

23.3.1.1 Características estáticas

La figura 23.2a presenta la función de transferencia estática de un sistema que realiza la conversión digital-analógica, denominado convertidor digital-analógico (CD/A). La figura 23.2b ofrece el esquema simplificado de un convertidor D/A con salida y referencia de tensión.

La línea continua en a) representa la función de transferencia correspondiente a un convertidor de un número infinito de bits, mientras que los puntos representan la de un convertidor que interpreta a su entrada un código unipolar binario de 3 bits (8 códigos). La salida se presenta normalizada a su valor máximo, correspondiente al código binario de entrada de mayor valor (1111...1).

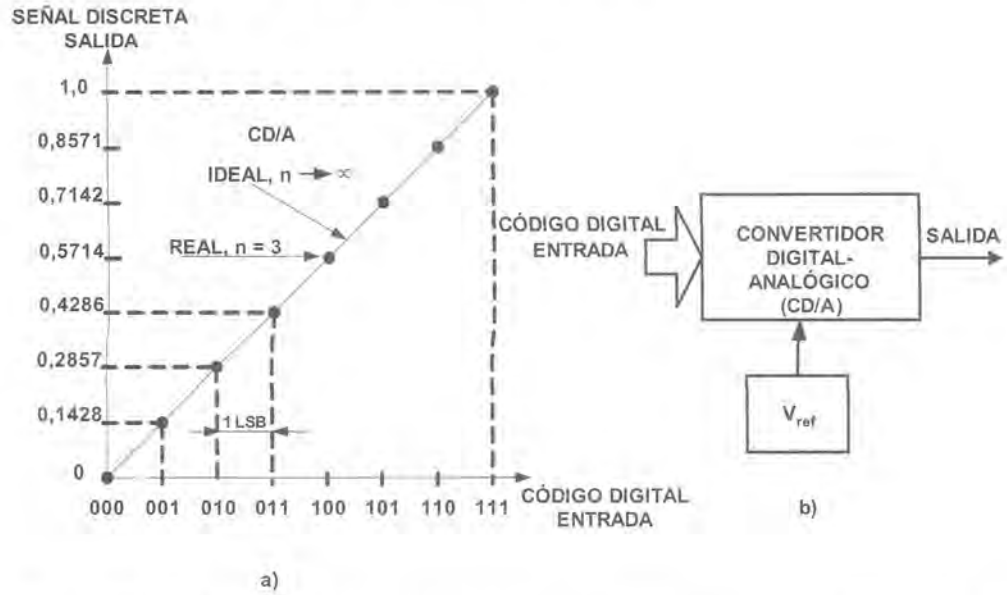


Figura 23.2. a) Función de transferencia de un sistema de conversión digital-analógica; b) convertidor digital-analógico.

Ambas características son **monótonas**, siendo esta una de las principales exigencias para un convertidor D/A. Si Q es el número total de códigos de entrada que interpreta un convertidor D/A y $Q = 2^n$, se suele decir que el convertidor tiene una **resolución** de n bits. Si para el código correspondiente al cero hay cero señal en su salida y para el de plena escala la salida se corresponde con la máxima, se dice que la ganancia es la unidad. Veamos a continuación los parámetros que caracterizan estáticamente a un convertidor D/A[23.7] y confirmemos qué hay de cierto en estas definiciones.

- **Resolución:** dos códigos de entrada adyacentes se diferencian en 1 LSB y esta diferencia es la **resolución** que puede expresarse en términos del peso porcentual de 1 LSB respecto al fondo de escala mediante:

$$\text{Resolución [\%]} = \frac{1}{2^n - 1} \cdot 100$$

Y, si n es muy grande, puede aproximarse a:

$$\text{Resolución [\%]} = \frac{1}{2^n} \cdot 100 \qquad (23.11)$$

- **Ganancia** (*Gain*): la ganancia de un convertidor D/A o **sensibilidad** es la pendiente de su característica transferencial estática.

Es la constante que relaciona el valor de la tensión o corriente de salida correspondiente al código de plena escala (V_{oFS}; I_{oFS}) con la magnitud de la referencia de tensión o de corriente utilizada en el circuito (V_{ref}; I_{ref}) para producir los elementos de la suma ponderada (V_{pond} o I_{pond}) de las expresiones (23.10). Para el caso de un convertidor D/A de salida en tensión que utiliza una referencia de tensión V_{ref} será:

$$G = \frac{V_{FS}}{V_{ref}} \qquad (23.12)$$

Muy a menudo se ajusta esta ganancia a la unidad, con lo que la salida correspondiente al código de plena escala coincide con la magnitud de la referencia de tensión o corriente utilizada en el convertidor.

- **Exactitud** (*Accuracy*): la exactitud de un convertidor D/A se refiere a la diferencia entre el valor que se obtiene en la salida para un determinado código de la entrada y el valor esperado.

Se expresa en términos de LSB. Por ejemplo, un convertidor D/A de 10 bits, de resolución 0,01% y de exactitud de ±½ LSB o de ±0,005%.

- **Exactitud relativa** (*Relative accuracy*): este parámetro se refiere a la diferencia entre el valor que se obtiene en la salida para un código de entrada y el valor esperado, luego de ajustar la característica transferencial estática a la ideal en los puntos de cero y plena escala. Se expresa en LSB o en por ciento.

23.3.1.2 Características dinámicas

La figura 23.3 presenta la salida de un convertidor D/A de salida en tensión, V_o al producirse un cambio del código digital de su entrada I a otro N. Un convertidor ideal ofrecería un cambio inmediato de la salida, mientras que uno real pasa por un transitorio para alcanzar el valor correspondiente al código N. La caracterización dinámica de un convertidor D/A se suele expresar mediante los siguientes parámetros:

Tiempo de establecimiento (*Settling Time*): es la duración del transitorio (t_s en la figura 23.3) desde que se produce el cambio de código hasta que la salida se encuentre dentro de un margen de error, por lo general de ½ LSB, expresado en términos de la señal de salida. Se representa por las siglas **ST**.

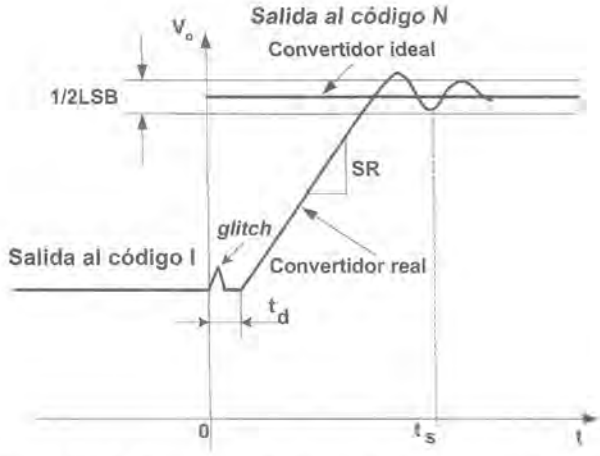


Figura 23.3. Respuesta dinámica de un convertidor D/A.

Este retardo está compuesto por un primer proceso de duración $\mathbf{t_d}$ durante el cual la conmutación en tiempos desiguales de los interruptores del circuito produce un efecto indeseable en la salida denominado "*glitch*"[3], un cambio de tensión con pendiente determinada por la rapidez de cambio SR del circuito y un proceso oscilatorio amortiguado alrededor del valor final de la tensión o corriente de salida. Su magnitud depende de la resolución, del tipo de circuito convertidor y de la tecnología de fabricación de sus elementos y, para convertidores comerciales en forma de circuito integrado, se encuentran desde decenas de nanosegundos hasta decenas de microsegundos.

Rapidez de cambio de la salida (***Slew Rate***): es el cociente entre la tensión o la corriente de plena escala (V_{oFSV} o I_{oFSV}) y el tiempo de establecimiento requerido para alcanzar el valor de plena escala partiendo del valor de cero. Se representa por las siglas **SR** y se expresa en V/s.

Frecuencia de conversión (*Conversion Rate*): es la frecuencia máxima a la que se puede cambiar el código de entrada obteniendo la salida correspondiente. Generalmente suele ser menor que el inverso del tiempo de establecimiento y se expresa en hertzios o en muestras por segundo (S/s). Por ejemplo, el convertidor D/A TL5632 de alta velocidad, 8 bits, $t_s = 15$ ns tiene una frecuencia de conversión de 60 MS/s (60 Megamuestras/s) y el inverso de $\mathbf{t_s}$ es 66,6 MHz.

EJERCICIO 23.6

Un convertidor D/A de 8 bits se utiliza para "reproducir" una rampa lineal.

a) Calcule el número máximo de pasos o escalones de tensión que tendrá la salida.

b) ¿Cuál es la pendiente máxima de la "rampa escalonada" que puede ser reproducida con la calidad correspondiente a la resolución de este convertidor?

SOLUCIÓN:

a) El número de pasos de tensión diferentes que tendrá la señal de salida depende de la resolución del convertidor. En este caso el convertidor es de código unipolar binario directo y, alimentado con 5 V, presenta la siguiente tabla de asociación código-tensión de salida V_o.

[3] La magnitud de este efecto se suele nominar como impulso *glitch*. Depende del tipo de convertidor y se expresa en términos de su energía (V-s). Su valor más crítico es cuando se produce la conmutación de todos los interruptores, es decir, entre códigos adyacentes correspondientes a media escala del convertidor (011...1 → 100...0).

Código de entrada	V_0, V
00000000	3,980
11111111	5,000

Para reproducir una rampa, esta partirá de la tensión correspondiente al código 00000000 (3,980 V) y llegará a su valor máximo con el código 11111111 (5,000 V) y en este margen de tensión existirán 2^8 pasos o escalones de tensión diferentes, incluyendo los extremos.

b) La mayor frecuencia a la que pueden cambiar los datos a la entrada de este convertidor está dada por su frecuencia de conversión máxima que es 60 MHz. Si se parte del código 00000000 y sucesivamente se va incrementando en 1 LSB hasta arribar al código 11111111 a la frecuencia de 60 MHz, se obtendrá a la salida la "rampa" de pendiente máxima $m_{máx}$ formada por $2^8 = 256$ escalones (valores discretos) de tensión de igual duración (si no consideramos los errores de linealidad) e igual al periodo de cambio, que es 0,0167 μs (1/60 MHz). Esta pendiente será:

$$m_{máx} = \frac{5,000 - 3,980}{256 \cdot (0,0167) \cdot 10^{-6}} = 239 \text{ V/ms}$$

23.3.1.3 Errores en los convertidores D/A

En los convertidores D/A se presentan un conjunto de errores dependientes de los componentes de sus circuitos y de sus derivas temporales y térmicas. Entre los errores fundamentales pueden citarse el **error de no-monotoneidad**, el **error de cero**, el **error de ganancia**, y los **errores de linealidad integral y diferencial**.

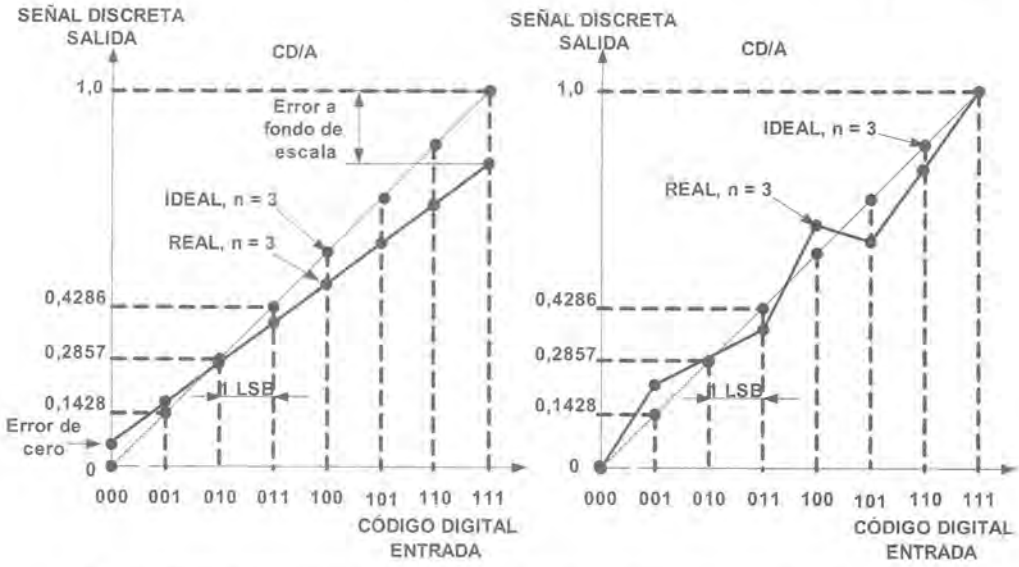


Figura 23.4. Ilustración de los errores fundamentales de un convertidor D/A.

Errores de cero, de plena escala y de ganancia

La figura 23.4 presenta la característica de transferencia estática con salida normalizada de un convertidor D/A ideal unipolar de 3 bits y de uno real con estos errores. En ambos casos, hemos unido mediante segmentos de recta los puntos que realmente conforman las características.

En la gráfica de la izquierda se presenta el **error de cero** o de *offset* a la vez que el **error a fondo de escala**, como las diferencias entre las salidas ideales y las reales que corresponden a los códigos del cero y de la plena escala, respectivamente. Estos errores se expresan en términos del LSB. La suma de los errores de cero y a fondo de escala es equivalente al **error de ganancia**, que se expresa en los mismos términos que sus componentes.

EJERCICIO 23.7

Un convertidor D/A de 8 bits de código unipolar binario directo tiene una fuente de referencia $V_{ref} = 10$ V. Su salida para el código 00000000 es 0,5 mV y para el código 11111111 es 9,9 V.

a) Exprese su resolución en tanto por ciento.

b) Calcule su ganancia real.

c) ¿Qué errores tiene de cero, a fondo de escala y de ganancia?

SOLUCIÓN:

a) Para n = 8 bits, %Resolución = 100%/(2^8-1) = 0,3922%. Si se aplica la expresión aproximada, se obtiene %Resolución$_a$ = 100%/(2^8) = 0,3906%. Obsérvese que la diferencia entre los valores de resolución calculados es muy pequeña (sólo de un 0,4%), por lo que se suele calcular para n ≥ 8 bits mediante la expresión aproximada.

b) La ganancia real de este convertidor será: G = (9,9 – 0,0005)/10 = 0,98995.

c) El peso en términos de la salida de 1 LSB será: 1 LSB = 10/(2^8–1) = 0,03922 V = 39,22 mV.

El error de cero será: 0,5 mV/39,22 mV/LSB = 0,0127 LSB

El error a fondo de escala será: (10 – 9,9) V/0,03922 V/LSB = 2,5497 LSB

El error de ganancia será: 2,5497 + 0,0127 = 2,5624 LSB

Errores de linealidad

En la gráfica de la derecha de la figura 23.4 se han exagerado las desviaciones de la característica real respecto a la ideal. Se puede observar que para el código 101 la salida es menor que para el código adyacente inferior (100). En esto consiste el **error de no-monotoneidad** que es un caso particular de error de no-linealidad diferencial.

Las salidas correspondientes a dos códigos de entrada adyacentes (si el convertidor es ideal) se diferencian en 1 LSB (expresado en términos de la salida). Cuando el convertidor experimenta errores de no-linealidad esto no se cumple.

El error de **no-linealidad diferencial** (***DNLE** Differential Nonlinearity Error*) es el módulo de la diferencia entre los valores de la salida correspondientes a dos códigos adyacentes menos 1 LSB y se expresa en términos de LSB. Así las cosas, para un CD/A cuya salida en tensión, se puede expresar como:

$$\text{Error de no-linealidad diferencial (DNLE)} = \frac{\left|V_{oJ+1} - V_{oJ}\right| - V_{oLSB}}{V_{oLSB}} \text{ [LSB]} \qquad (23.13)$$

Si para algún código J de entrada la diferencia (V_{oJ+1}–V_{oJ}) es negativa, se ha producido un error de **no-monotoneidad** en el código J+1. Hoy día una parte importante de los convertidores D/A comerciales inte-

grados garantizan la monotoneidad de la característica transferencial estática en un amplio margen de temperatura de trabajo.

El error de **no-linealidad integral** (*INLE Integral Nonlinearity Error*) es la suma de los errores de no-linealidad que posee el convertidor en un intervalo de códigos de la entrada, y puede expresarse como:

$$\text{Error de no-linealidad integral (INLE)} = \frac{1}{V_{oLSB}} \cdot \sum_{J=0}^{N} \left| V_{oJ+1} - V_{oJ} \right| - V_{oLSB} \quad [\text{LSB}] \qquad (23.14)$$

Como **error de no-linealidad** suele identificarse también por los fabricantes al mayor valor del error de no-linealidad diferencial del convertidor en todo el margen de su entrada.

Las **derivas térmicas** de estos parámetros introducen también errores. Por ejemplo, el **coeficiente térmico de la ganancia** de un convertidor D/A expresa el cambio en el error de ganancia al variar la temperatura en un determinado margen y se expresa en tanto por ciento por cada grado centígrado, partes por millón por grado centígrado o como fracción de LSB por grado centígrado. También las derivas térmicas de los errores de **linealidad**, de **cero** y de **plena escala**, se expresan en los mismos términos.

Particular importancia reviste la no-igualdad del error de no-linealidad diferencial debido a que la contribución a la tensión o corriente de salida de cada bit del código de entrada no es independiente del estado de los restantes bits. A esto se le denomina **error de superposición** [23.4] y sólo se minimiza utilizando redes apropiadas en los convertidores.

EJERCICIO 23.8

Un convertidor D/A de 16 bits con salida en tensión, trabajando con código de entrada unipolar binario presenta (versión K, 0 ÷ 70°C) los siguientes parámetros:

- *Exactitud relativa a 25°C: ±4 LSB.*
- *Error de no-linealidad diferencial: ±0,5 LSB.*
- *Error de no-linealidad integral: ±2 LSB.*
- *Error de ganancia a 25°: ±6 LSB.*
- *Error de cero a 25°:±3 LSB.*
- *Coeficiente térmico del error de ganancia (Gain TC): ±2 ppm FSR/°C.*
- *Coeficiente térmico del error de cero (Offset TC): ±2 ppm FSR/°C.*

a) *Calcule la magnitud de los errores de ganancia y de cero a 70°C si la salida se ofrece en el margen de 0 ÷ 10 V.*

b) *Para el código de entrada de media escala 1000 0000 0000 0000, calcule la salida ideal y la salida real a 25°C si el margen de tensión de salida es 0 ÷ 10 V.*

Nota: *FSR es el margen estático de salida del convertidor.*

SOLUCIÓN:

a) Si el convertidor tiene un margen de tensión de salida 0 ÷ 10 V, el peso en voltios de 1 LSB será:

$$1 \text{ LSB} = (10 - 0)/2^{16} = 153 \ \mu\text{V}$$

Sus coeficientes térmicos de ganancia y *offset* serán:

$$\textit{Gain} \text{ TC} = \textit{Offset} \text{ TC} = \pm 2 \text{ ppm FSR/°C} = \pm 2 \cdot 10^{-6} \cdot 10/^{\circ}\text{C} = \pm 20 \ \mu\text{V/°C}$$
$$= \pm 20 \ \mu\text{V/°C}/153 \ \mu\text{V/LSB} = \pm 0{,}13 \text{ LSB/°C}$$

El error de cero a 70ºC será: ±3 LSB ± 0,13 LSB/ºC·45ºC = ± 8,85 LSB.

El error de ganancia a 70ºC será: ±6 LSB ± 0,13 LSB/ºC·45ºC = ±11,85 LSB

b) Si el convertidor no tuviera errores, la salida para el código de media escala 8000 H sería: $V_{oI\ (8000H)} = 10 \cdot 2^{15}/(2^{16}-1) = 5{,}0000763$ V.

El fabricante ofrece un valor para la exactitud relativa del convertidor y es el que debe ser utilizado, ya que el circuito posee ajuste de los errores de cero y de ganancia con lo que se ajusta el error a fondo de escala. Así las cosas, el valor real de la salida para esta combinación será: $V_{oR\ (8000H)} = 5{,}0000763 \pm 4 \cdot 153 \cdot 10^{-6}$ V, o se encontrará entre 4,9994643 V y 5,0006868 V.

23.3.2 TIPOS DE CONVERTIDORES D/A

La suma ponderada de corrientes o tensiones expresada en (23.10) se realiza en la práctica utilizando fuentes de referencia de tensión y/o fuentes de corriente y una red de interruptores, resistencias y/o condensadores. La conversión digital-analógica generalmente consiste en **obtener una tensión o corriente mediante la división o la multiplicación de una tensión, corriente o carga eléctrica de referencia por una cantidad dependiente del código de entrada** dando lugar a una gran variedad de circuitos convertidores de los que nos referiremos a los más significativos.

23.3.2.1 Convertidores de elementos ponderados en binario

Convertidor de resistencias ponderadas. La figura 23.5 presenta el diagrama básico de este convertidor que consta de tres partes: la fuente de referencia V_{ref}, la red de resistencias ponderadas (2R, 4R...2^nR) en número igual al número de bits de información del código de entrada y el convertidor corriente-tensión (I/V), si se desea que la salida sea en tensión.

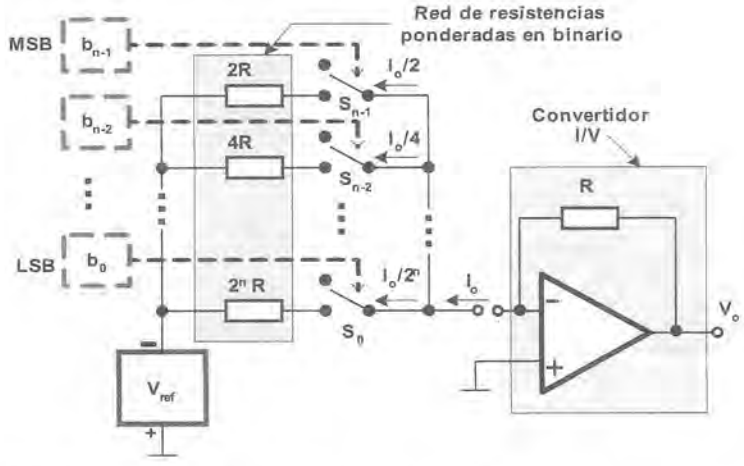


Figura 23.5. Esquema de principio del convertidor D/A de resistencias ponderadas

Cuando todos los bits del código están en 1, todos los interruptores S se encuentran cerrados, y la corriente de salida I_o será máxima (plena escala, I_{oFSV}) e igual a $I_{oFSV} = V_{ref}/R$. Cuando b_{n-1} (el MSB de información) está en 1, la corriente que aporta esta rama a la salida es *la mitad* de la máxima $I_{oFSV}/2 = V_{ref}/2R$. Así

las cosas, b_{n-2} aportará la *cuarta parte* ($V_{ref}/2^2.R$) y b_0 (el LSB) sólo aportará $V_{ref}/(2^n.R)$. En forma general, la salida I_o se expresará como:

$$-I_o = \frac{V_{ref}}{2^n R} \cdot \sum_{i=0}^{n-1} 2^i \cdot bi \text{, donde } bi \in [0,1] \qquad (23.15)$$

Para obtener una salida en tensión se conecta un convertidor I/V que aumenta el tiempo de establecimiento t_s por el atraso que este incorpora al sistema y la salida V_o será:

$$V_o = \frac{V_{ref}}{2^n} \cdot \sum_{i=0}^{n-1} 2^i \cdot bi \text{, donde } bi \in [0,1] \qquad (23.16)$$

Obsérvese que V_o y V_{ref} tienen polaridades opuestas; en el caso de utilizar el código bipolar binario con signo, el bit correspondiente al signo sirve para determinar la polaridad de V_{ref} y con ella la de la salida.

El problema fundamental de este circuito es que necesita resistencias de precisión muy diferentes entre sí, además de que no evita el error de superposición. Por ejemplo, si R = 5 kΩ y el convertidor es de n = 12 bits, la resistencia asociada al LSB será de 4096 R = ¡¡20,48 MΩ!! En un circuito integrado tal valor de resistencia resulta muy difícil de lograr. Además, en este convertidor la impedancia de la red es variable con el código de entrada con lo que si alguna resistencia cambia de valor ese error afectaría a todos los códigos en que el bit asociado a esta sea 1. Debido a ello, este sistema sólo se utilizan para arquitecturas combinadas en forma de elemento de 4 bits o utilizando una cascada de resistencias ponderadas, como la presentada en la figura 23.6, con la que se construye un convertidor de 12 bits con sólo 4 tipos de resistencias, además de las del atenuador de corriente de relación 15:1, pero su linealidad integral depende en gran medida de la calidad de los atenuadores.

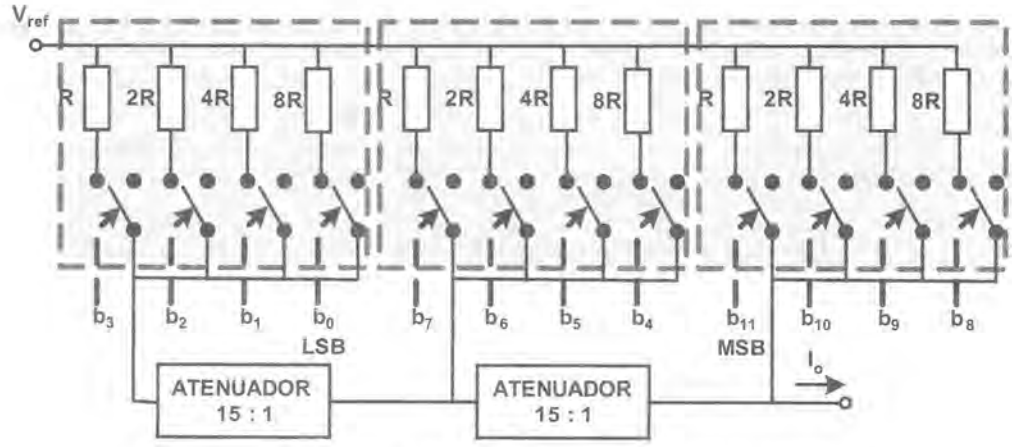


Figura 23.6. Cascada de resistencias ponderadas en binario.

Convertidor de fuentes de corriente ponderadas en binario. Si la red de resistencias ponderadas de la figura 23.5 es sustituida por una red de fuentes de corriente ponderadas en binario (figura 23.7) se obtiene idéntico resultado, pero con las ventajas de que la corriente por los interruptores no varía al cambiar el código de entrada, y que a la salida se obtienen corrientes complementarias[4].

[4] La suma de estas corrientes es constante e igual a la corriente de plena escala $I_{oFSV} = V_{ref}/R$.

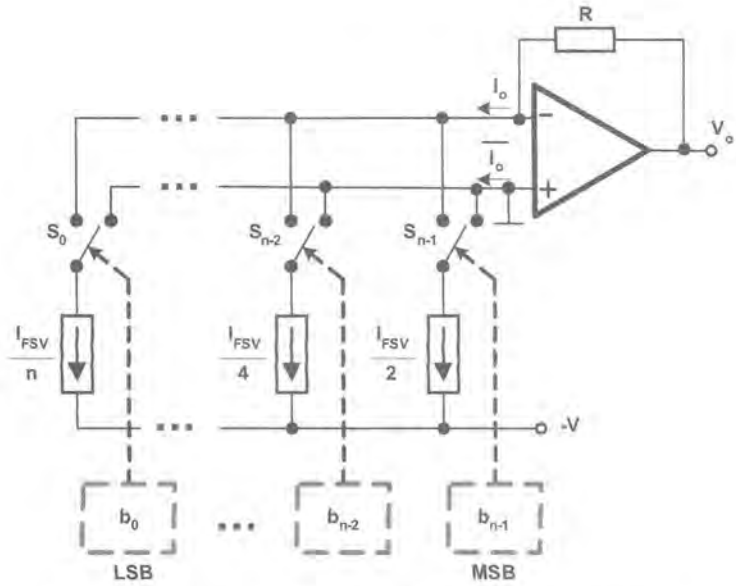


Figura 23.7. Convertidor de fuentes de corriente ponderadas en binario.

Si las fuentes y los interruptores son construidos con transistores bipolares, los resultados serán muy buenos. Como desventaja fundamental puede apuntarse que el área de silicio requerida en el circuito integrado para los transistores de las fuentes de corriente aumenta en proporción directa con 2^{n-1} y, para un número superior a n = 8, la más pequeña de las corrientes puede ser enmascarada por el ruido en el circuito. Una solución resulta de combinar fuentes de corriente de sólo cuatro valores con atenuadores, como se muestra en la figura 23.8 para un convertidor de 8 bits.

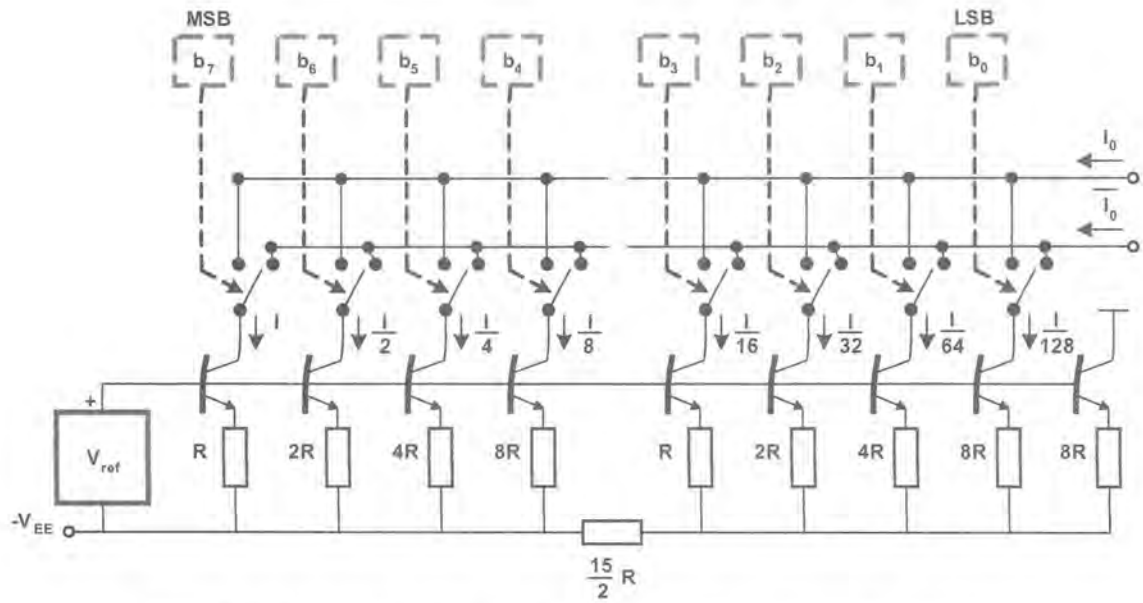


Figura 23.8. Red de fuentes de corriente ponderadas en secciones de 4 bits.

Convertidor de capacidades ponderadas en binario. La figura 23.9 presenta un esquema simplificado de este convertidor, basado en el principio de la redistribución de carga entre condensadores conectados en serie y/o en paralelo.

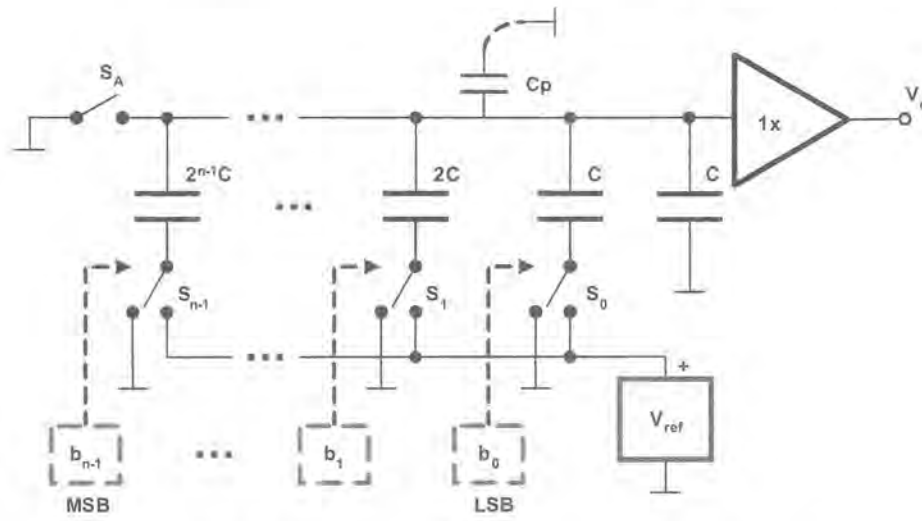


Figura 23.9. Convertidor de capacidades ponderadas en binario.

Su funcionamiento requiere dos fases marcadas por dos ciclos de un reloj: **la de inicialización**, en la que con el interruptor S_A cerrado y los interruptores de ramas del circuito ($S_0S_1...S_{n-1}$) en la posición de masa se garantiza que todos los condensadores estén descargados, y **la de redistribución de la carga**, que comienza con abrir S_A a la vez que los interruptores ($S_0S_1...S_{n-1}$) se colocan a V_{ref} si el bit asociado es "1" o se quedan a masa si es "0". Así las cosas, se forma un divisor capacitivo que da como resultado, luego del transitorio de redistribución de las cargas en todos los condensadores, que la tensión de salida sea:

$$V_o = \left[\frac{2^n \cdot C}{2^n \cdot C + Cp}\right] \cdot \frac{V_{ref}}{2^n} \cdot \sum_{i=0}^{n-1} 2^i \cdot b_i \text{, donde } b_i \in [0,1] \qquad (23.17)$$

Este circuito tiene la ventaja de que puede construirse monolíticamente con tecnología CMOS de forma fácil lo que implica muy bajo consumo de energía (sólo las pérdidas en la conmutación de los interruptores), pero presenta las desventajas de un apreciable error de ganancia debido a la capacidad parásita Cp del punto común del conjunto de condensadores, valores muy diferentes de capacidades si n es grande que dificulta lograr un pequeño error de linealidad y un tiempo de establecimiento mayor que otras técnicas, que se incrementa con n al aumentar el número de ramas y el valor de sus capacidades.

23.3.2.2 Convertidores de código de termómetro

Ciertos convertidores D/A conforman su salida analógica mediante la suma de "unidades" de tensión o corriente iguales y de magnitud igual al peso del LSB. Ello implica que si el convertidor es de n bits dispondrá de $(2^n - 1)$ "unidades". Para el código correspondiente a la plena escala, todas las unidades son seleccionadas y para el del cero, ninguna. El código de selección de las unidades que responde a esta asociación tan simple recibe el nombre de **código de termómetro**, por su similitud con el funcionamiento de un termómetro de columna de mercurio. La tabla 23.2 ilustra la correspondencia entre los códigos decimal, binario y de termómetro para n = 3 bits. Los convertidores D/A que responden a este código presentan ventajas frente a

su contrapartida de elementos ponderados en binario tales como menores errores de linealidad diferencial, garantizan la monotoneidad, reducen el ruido *"glitch"* y aunque aumentan en número la cantidad de componentes, no incrementan en gran medida el área como circuito integrado monolítico para n moderadamente grande.

	BINARIO			DE TERMÓMETRO						
DECIMAL	b_0	b_1	b_2	d_1	d_2	d_3	d_4	d_5	d_6	d_7
0	0	0	0	0	0	0	0	0	0	0
1	0	0	1	0	0	0	0	0	0	1
2	0	1	0	0	0	0	0	0	1	1
3	0	1	1	0	0	0	0	1	1	1
4	1	0	0	0	0	0	1	1	1	1
5	1	0	1	0	0	1	1	1	1	1
6	1	1	0	0	1	1	1	1	1	1
7	1	1	1	1	1	1	1	1	1	1

Tabla 23.2. Código de termómetro.

Convertidor de cadena de resistencias. La figura 23.10 presenta el esquema de principio de este convertidor para 3 bits con el propósito de apreciar fácilmente la arquitectura del decodificador.

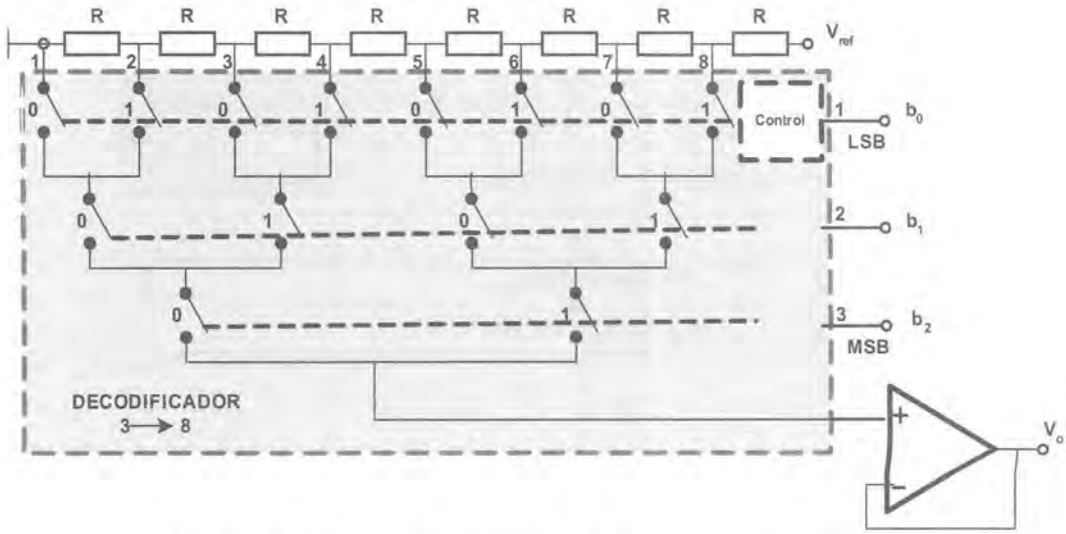


Figura 23.10. Convertidor de cadena de resistencias para n = 3 bits.

Si las 2^n resistencias son iguales entre sí y estables, este convertidor exhibirá monotoneidad, por lo que también recibe el nombre de **inherentemente monótono**. La tensión de salida será:

$$V_o = \frac{V_{ref}}{2^n} \cdot \left[b_{n-1} b_{n-2} \ldots b_0\right]_{DEC} \tag{23.18}$$

donde n es el número de bits del convertidor y $[b_{n-1}b_{n-2}...b_0]_{DEC}$ es el código aplicado al convertidor expresado en decimal. La tensión a fondo de escala será:

$$V_{oFS} = \frac{V_{ref}}{2^n} \cdot \left(2^n - 1\right)$$

Este convertidor se caracteriza por ofrecer tiempos de establecimiento muy pequeños, pero su linealidad integral depende del equilibrio existente entre sus resistencias. Su principal desventaja es el número de componentes: el número total de resistencias es 2^n y el de interruptores $(2^{n+1} - 2)$ por lo que para **n** muy grande requiere de una gran área de silicio lo que hace que sólo se fabrique para $n \leq 8$ bits.

Convertidor conmutador de corrientes (*current-switch DAC*). Una versión de convertidor inherentemente monótono muy rápido es el denominado de conmutación de corrientes, donde 2^n-1 fuentes de corriente idénticas **I** hacen llegar su corriente a la salida como se presenta esquemáticamente en la figura 23.11 para tecnología MOS. El decodificador de n entradas y 2^n salidas de código de termómetro activa las fuentes de corriente de manera tal que, si el código de entrada es 00...0, ninguna se encuentra activada e $I_o = 0$ y, con la combinación 11...1, todas están activadas e $I_o = I_{oFSV} = (2^n - 1)\cdot\mathbf{I}$. Su principal desventaja es que la linealidad integral depende de la igualdad de las fuentes de corriente y que, como circuito integrado, puede llegar a ser inviable si n es muy grande, presentando un gran consumo.

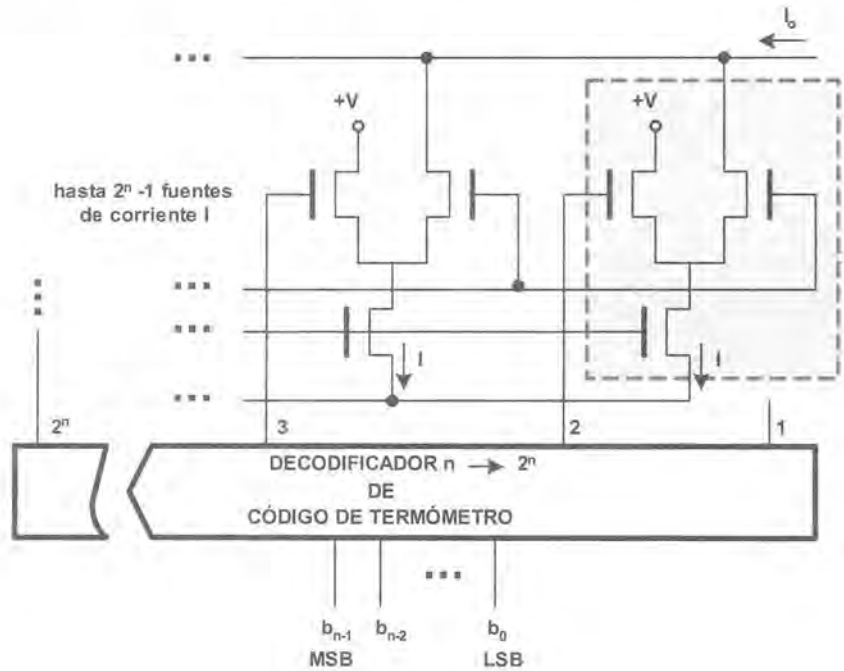


Figura 23.11. Convertidor de corrientes conmutadas con tecnología MOS.

Convertidor de capacidades conmutadas (*switched-capacitor DAC*). La versión con capacidades del convertidor de código de termómetro se presenta en la figura 23.12. Como el de capacidades ponderadas, este circuito necesita dos ciclos de trabajo: **inicialización**, donde todos los condensadores iguales de ramas C son conectados a V_{ref} y el condensador C_f se mantiene descargado por el cierre de los interruptores S_A. **Transferencia de carga**, donde los S_A se abren, se cierran los interruptores de rama ($S_1S_2...S_2{}^n{}_{-1}$) en los que la salida del decodificador es "1", la carga de todos los condensadores conectados a la entrada del integrador es transferida al condensador C_f y los condensadores que no han sido conectados a la entrada del integrador durante este ciclo deben descargarse. Si el valor en decimal del código binario de entrada es **m**. la tensión de salida (con S_B cerrado) será:

$$V_o = V_{ref} \cdot \frac{m \cdot C}{C_f} \tag{23.19}$$

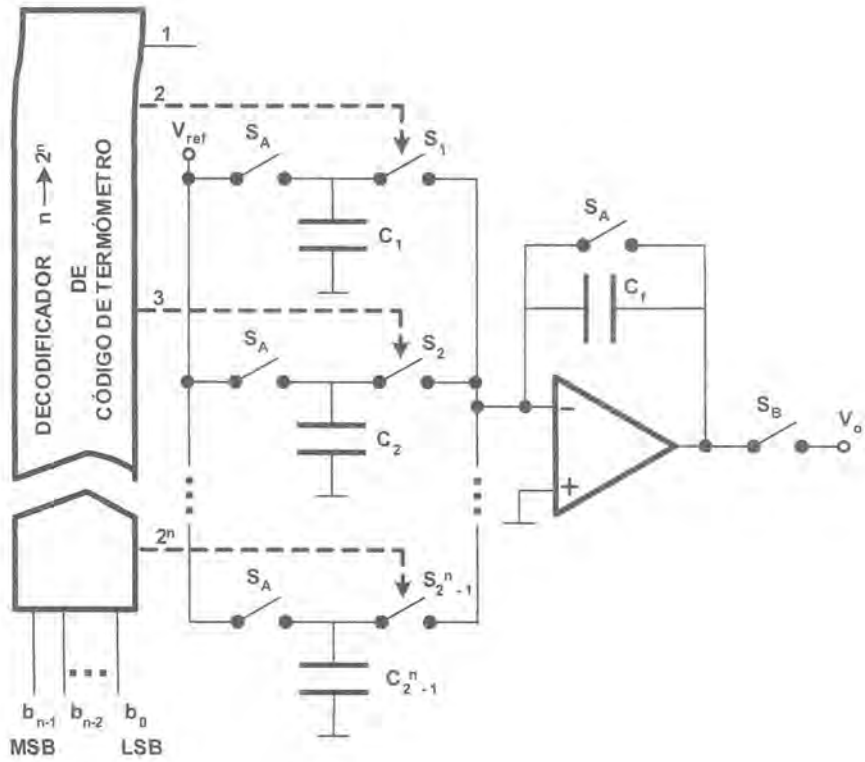


Figura 23.12. Convertidor de capacidades conmutadas.

Si se desea que a fondo de escala la tensión de salida sea muy parecida a V_{ref}, entonces $C_f = 2^n \cdot C$.

Este convertidor puede construirse para n hasta 10 bits y trabajar a muy altas frecuencias de conversión si C es pequeña, pero para valores inferiores a 1 pF resulta muy difícil lograr la igualdad de todos los condensadores lo que introduce un importante error de linealidad.

23.3.2.3 Convertidores de redes escalera (*ladder*)

Una red escalera está formada por elementos pasivos de sólo dos valores diferentes, en la que su impedancia equivalente resulta bastante independiente del código aplicado a la entrada, minimizando los errores por superposición y en la que no resulta muy crítico que los componentes mantengan su valor exacto.

Este tipo de redes puede ser implementado con resistencias o con condensadores y tienen la ventaja sobre los de elementos ponderados de poseer menor número de componentes y con sólo dos valores, presentando menores errores de linealidad.

Convertidor de red R-2R. La figura 23.13a presenta la red escalera correspondiente a un convertidor D/A de 4 bits en la que si los valores de las resistencias están en proporción 2:1 la resistencia equivalente de la red que se ve a la izquierda de cada resistencia es R. El número de elementos de la red escalera es 2n-1, donde n es el número de bits del código de entrada.

El circuito equivalente Thévenin de cualquiera de las ramas asociadas a los interruptores S, es una fuente de valor $V_{ref}/2$ en serie con una resistencia R (figura 23.13b), por lo que $I = V_{ref}/2R$ es la corriente que circula por cada interruptor cuando está en la posición "1" y el resto en la posición "0".

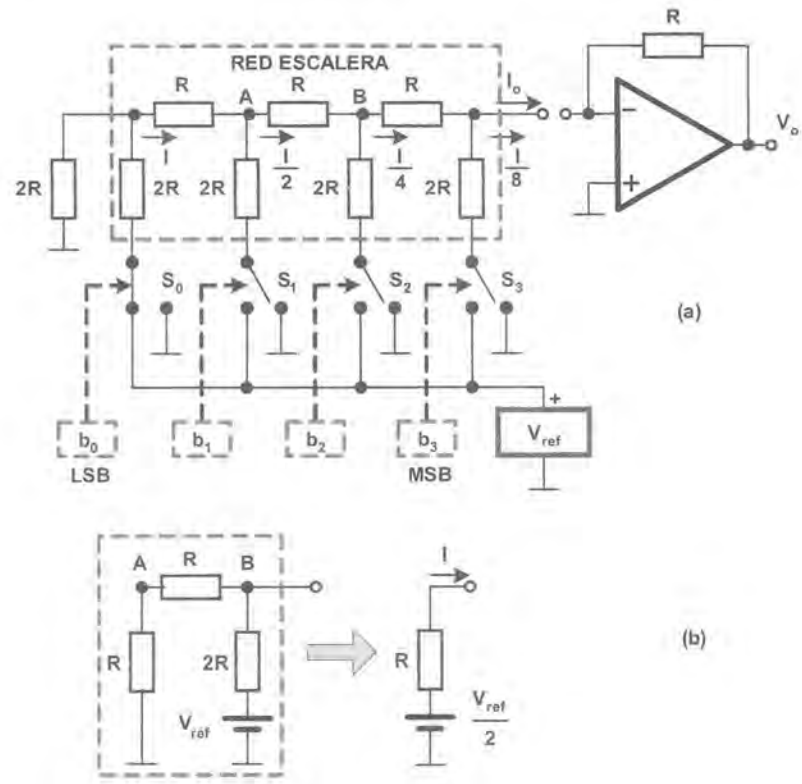


Figura 23.13. a) Esquema de principio del convertidor de red R-2R, b) circuito equivalente de la red desde un nodo.

El aporte que cada rama realiza a la corriente de salida I_o será el siguiente:

- El aporte de la rama asociada al MSB será I.
- El aporte de la rama asociada al LSB (figura 23.13a) será de sólo $I/8 = I/2^3$.

Así las cosas, la corriente de salida I_o, y la tensión V_o si se utiliza el convertidor I/V, se podrán expresar como:

$$I_o = \frac{V_{ref}}{2^n \cdot R} \cdot \sum_{i=0}^{n-1} 2^i \cdot b_i \quad , \text{donde } b_i \in [0,1]$$

$$V_o = -\frac{V_{ref}}{2^n} \cdot \sum_{i=0}^{n-1} 2^i \cdot b_i \quad , \text{donde } b_i \in [0,1] \qquad (23.20)$$

y la salida de tensión a fondo de escala será:

$$V_{oFS} = -V_{ref} \cdot \frac{2^n - 1}{2^n} \qquad (23.21)$$

Este circuito tiene la ventaja de poder ser implementado monolíticamente para un número grande de bits, sólo posee dos valores diferentes de resistencias, necesita interruptores de muy baja resistencia de conducción (están en serie con una resistencia equivalente de valor R) y de muy baja tensión de desviación. Una desventaja importante de este circuito es su limitación para frecuencias de conversión altas debido a que

cuando un bit cambia, la corriente por la resistencia 2R de esa rama se invierte, lo que provoca un transitorio de duración dependiente de las capacidades asociadas a estas resistencias.

Convertidor de red R-2R invertida. La figura 23.14 presenta la red R – 2R para este convertidor D/A. Difiere de la anterior en que la fuente de tensión de referencia no alimenta a las ramas de resistencias-interruptores, sino al extremo de la red escalera desde donde la impedancia de la red es independiente del código de la entrada.

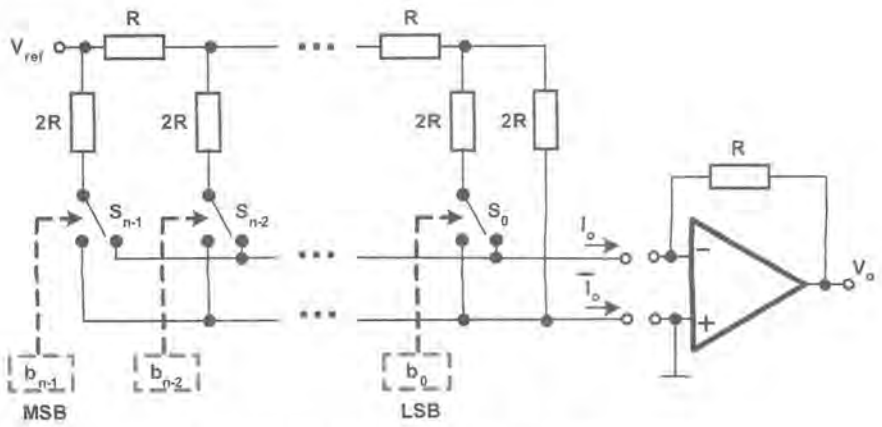


Figura 23.14. Convertidor de red R-2R invertida.

Al cambiar los bits del código de entrada, la corriente por cada rama sigue circulando en la misma dirección por lo que se reduce el transitorio asociado a ese cambio y la corriente por la red es siempre la misma. Este circuito permite tener disponibles corrientes de salida I_0 complementarias, por lo que también es conocida como red escalera en modo de corriente. Al conectar en la salida un convertidor I/V, los interruptores siempre conmutan a masa por lo que la tensión a la que están sometidos es muy baja, no afectando su conmutación a la fuente de referencia. Si en calidad de V_{ref} se utiliza una tensión externa variable, obtenemos un convertidor D/A multiplicativo sin influencia mutua entre las variables a multiplicar (tensión de referencia y código), siendo esta la arquitectura más utilizada en los convertidores multiplicativos. Sin embargo, su impedancia de salida es variable con el código de entrada, lo que puede afectar al trabajo del convertidor I/V. Es la configuración más popular realizada en tecnología CMOS, las resistencias se fabrican difundidas, implantadas o por capas delgadas, dependiendo de la resolución requerida.

Convertidor de red C-2C. Si en la figura 23.13a sustituimos R por 2C y 2R por C, obtenemos el circuito de este convertidor. Mediante un análisis similar al realizado en la red R-2R se puede llegar a los mismos resultados, pero teniendo en cuenta que deben existir dos ciclos, uno de inicialización que garantiza que todos los condensadores se descarguen y otro en el que la carga de los condensadores de rama es transferida al condensador del integrador. Este convertidor ofrece menores tiempos de establecimiento que el de capacidades ponderadas en forma de circuito monolítico con tecnología CMOS siendo uno de los de más bajo consumo.

23.3.2.4 Convertidores de tensiones o corrientes segmentadas

Esta arquitectura es una de las más difundidas en los convertidores D/A de más de 8 bits ya que ofrece una solución de pequeños tiempos de establecimiento con una relativamente pequeña cantidad de componentes, además de reducir la energía del *glitch*, mejorar el comportamiento dinámico del convertidor y su linealidad. En esta arquitectura, el código de entrada es segmentado generalmente en dos partes: una parte con los bits

más significativos se aplica a un convertidor inherentemente monótono y la otra se aplica a otro convertidor (elementos ponderados en binario, red escalera, etc.). Ambos convertidores aportan sus salidas analógicas para conformar la salida del sistema.

Convertidor de tensiones segmentadas. La figura 23.15 presenta los esquemas simplificados de dos convertidores de tensiones segmentadas para n = 6 bits. En la figura a) se presenta el conocido como divisor de Kelvin-Varley, en el que los tres bits más significativos deciden la posición del selector doble A en alguna de las 2^3 resistencias iguales R_1 del divisor de Kelvin, mientras que los menos significativos deciden la posición del cursor B en el divisor de Varley, de 2^3 resistencias iguales R_2. En la figura b) se presenta el divisor de Kelvin seguido de un convertidor R-2R para los tres bits menos significativos. En ambos casos, la tensión a través de la resistencia seleccionada en el divisor de Kelvin es subdividida por el convertidor de la salida, acorde al peso de los bits menos significativos.

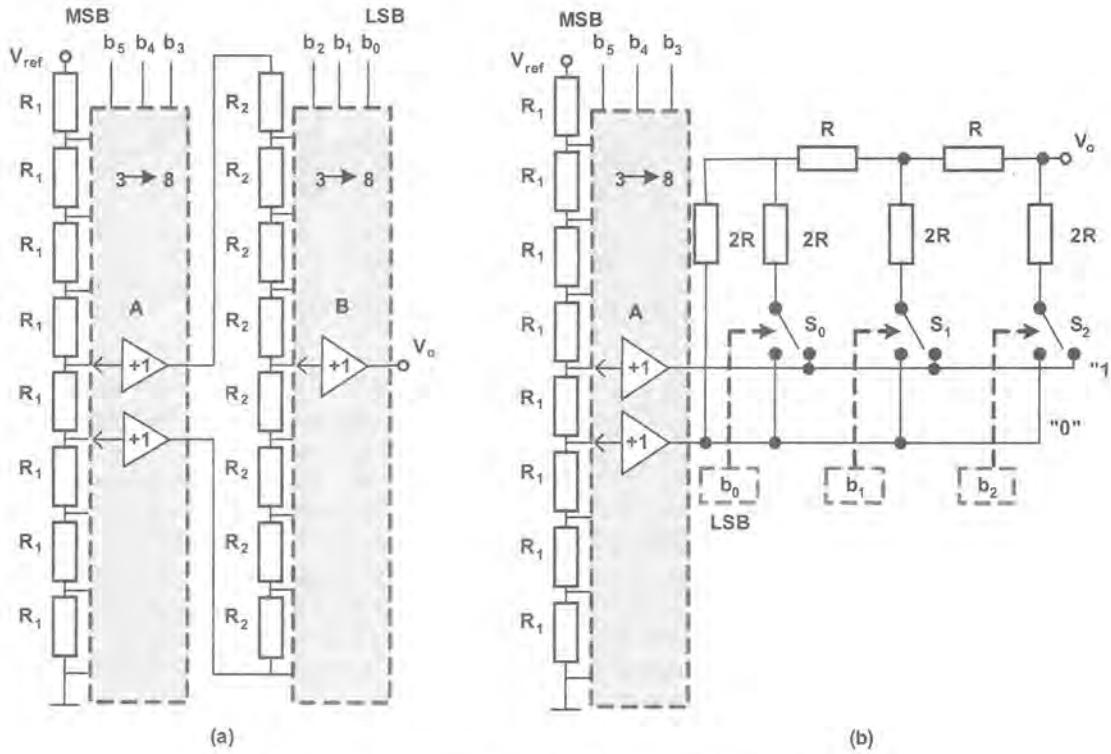


Figura 23.15. Convertidores de tensiones segmentadas.

El primero tiene la desventaja sobre el segundo de poseer un mayor número de resistencias, con lo que su consumo de energía es mayor, pero es más rápido. Por ejemplo, los convertidores:

AD569, 16 bits, BiCMOS, ST = 3 μs, arquitectura de Kelvin-Varley (8 bits y 8 bits), P_W = 1000 mW
AD7846, 16 bits, LC^2MOS, ST = 9 μs, arquitectura de Kelvin-R-2R (4 bits y 12 bits), P_W = 100 mW

Ambos son monótonos, pero el primero presenta un mayor error de no-linealidad (±0,02% frente a ±0,003%). Sus hojas de características se encuentran en el CD.

Convertidor de corrientes segmentadas. La figura 23.16 presenta la arquitectura de este convertidor para 8 bits, en la que se ha segmentado en partes iguales el código de entrada.

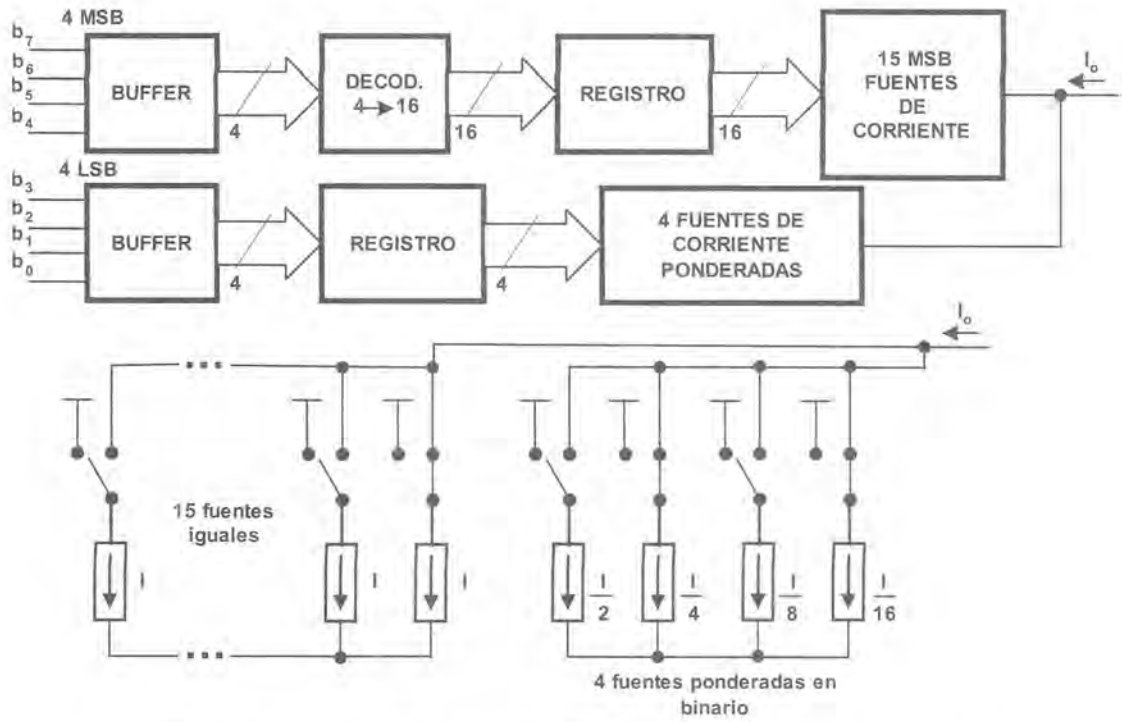


Figura 23.16. Arquitectura del convertidor de corrientes segmentadas para n = 8 bits.

Utilizando la parte más significativa ($b_7b_6b_5b_4$) se conecta a la salida un número de fuentes de corriente I igual al peso decimal que tenga el código, mientras que, con la parte menos significativa ($b_3b_2b_1b_0$), se decide la fracción de I que aportará este *nibble*. Es un convertidor muy rápido, no precisa muchas fuentes de corriente de diferentes valores y es una de las arquitecturas que más se aplica en forma de circuito integrado. Por ejemplo, el CD/A AD908, 8 bits, 165MS/s, $ST_{0,1\%} = 30$ ns presenta sólo 3 pV-s de *glitch*, pero es capaz de ofrecer una salida a fondo de escala de hasta $I_{oFSV} = 20$ mA a través de muy alta impedancia ($Z_0 = 200$ kΩ). Las hojas de características se encuentran en el CD.

Otro ejemplo de estos convertidores es el AD1866, también de Analog Devices, que presenta una arquitectura de 2 canales, entrada serie, 16 bits a una frecuencia máxima de reloj de 14 MHz, convertidor D/A de corrientes segmentadas: 8 bits en fuentes de corriente y 8 bits en red R-2R invertida, salida de tensión en ±1 V, bajo consumo (50 mW) y buenas prestaciones para aplicaciones de audio para computadores personales, equipos de CD portátiles y de coches. La hoja de características del fabricante se encuentra en el CD.

23.3.2.5 Convertidores Sigma-Delta

El audio digital profesional requiere convertidores D/A con resoluciones elevadas, muy buena linealidad y bajo coste. Los convertidores de arquitectura segmentada pueden lograr resoluciones de hasta 18 bits, pero los errores de linealidad son elevados, sobre todo en las transiciones del segmento que trabaja con los bits más significativos del código. Una solución a este problema la aportan los convertidores sigma-delta, cuya arquitectura general se presenta en la figura 23.17.

Este convertidor consiste en un filtro digital de interpolación (un circuito digital que acepta datos a una baja frecuencia, inserta ceros a una alta frecuencia, y aplica un algoritmo de filtrado digital que da como resultado una señal digital de más alta frecuencia), un modulador Σ-Δ (que actúa como un filtro paso bajo

para la señal y, a la vez, como un filtro paso alto para el ruido de cuantificación, y que convierte el dato resultante en una secuencia de bits de alta frecuencia) y un convertidor D/A de 1 bit (que consiste en un simple interruptor que coloca a su salida $+V_{ref}$ si la entrada es "1", o $-V_{ref}$ si la entrada es "0"). Por último, la salida es filtrada para obtener la señal analógica correspondiente al código de entrada [23.5][23.6].

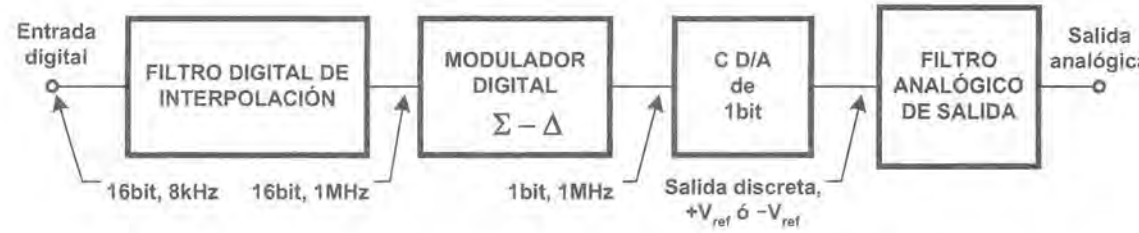


Figura 23.17. Arquitectura general de un convertidor sigma –delta (ΣΔ CD/A).

El hecho de utilizar técnicas digitales hace que su funcionamiento sea muy independiente de la temperatura; a la vez, el uso de un convertidor D/A de 1 bit hace que el sistema sea monótono y, en teoría, este convertidor es perfectamente lineal y genera muy bajo ruido en la misma medida en que las técnicas utilizadas logren pulsos de reloj de muy alta frecuencia y muy bajo ruido de fase y, a la vez, separen el ruido digital creado del espectro de la señal. Un ejemplo de CD/A de bajo coste, entrada serie, 16/20 bits, monótono, bajo consumo (2,5 mW), $ST_{0,003\%} = 15$ ms para 20 bits y con un error de linealidad máximo de ±0,0015% que con encapsulado SSOP de 16 pines y sólo dos condensadores externos para el filtro analógico de capacidades conmutadas, es el DAC1220, también de Analog Devices. Incorpora un microcontrolador que realiza las funciones de interpolación digital y corrección automática de los errores de ganancia y *offset*, que lo hacen ideal para aplicaciones de control en bucle cerrado en control de procesos de variables de amplio espectro de frecuencias y en instrumentación electrónica de alta resolución, así como en instrumentación portátil de alta calidad. Las hojas de características de este convertidor se encuentran en el CD.

23.3.3 ESTRUCTURAS DE LA ENTRADA EN LOS CIRCUITOS INTEGRADOS. CIRCUITOS MÚLTIPLES

23.3.3.1 Estructuras de la entrada

Los códigos a convertir en una señal analógica pueden llegar al convertidor básicamente en dos formas:

- **En serie** (secuencia de unos y ceros del código y algunos bits de control al inicio y/o al final de la cadena, a una frecuencia determinada).
- **En paralelo** (todos los bits del código se aplican simultáneamente al convertidor).

La duración de la carga del dato por parte del convertidor es menor si el dato se encuentra en paralelo, pero el circuito requiere tantos pines como bits para su entrada y esto, si n es grande, hace que el circuito integrado ocupe físicamente mucho espacio. Muchos convertidores de entrada paralelo realizan la carga por bytes o por *nibbles*, como una solución de compromiso entre tiempo y número de pines. Por otro lado, la entrada serie además de minimizar el espacio del circuito reduce el ruido capacitivo acoplado desde las líneas de datos hacia la salida analógica, lo que hace que, en la misma medida que se desarrollan convertidores de tiempos de establecimiento menores, se prefiere la entrada serie. Además, una gran parte de los procesadores modernos poseen interfaces de comunicación serie lo que facilita la compatibilidad de la comunicación procesador-convertidor.

Existe una gran variedad de estructuras de la entrada de datos de los convertidores D/A, mas algo común es la existencia de entradas doblemente almacenadas (*double-buffered*), que consiste en una cascada de dos "almacenadores" del código de la entrada. El primero recibe el dato en cualesquiera de las formas existentes (serie, paralelo, por bytes, etc.) y lo prepara para pasarlo al segundo en forma paralelo en el momento en que ya esté listo. Esto permite, entre otras cosas:

- Que la salida analógica se mantenga en el valor correspondiente al código anterior sin alteraciones durante la carga del siguiente código,
- Pasar datos a diferentes convertidores D/A y hacer que las señales analógicas correspondientes a ellos comiencen a estar presentes en las salidas simultáneamente, al dar la orden de convertir el dato a los segundos almacenadores a la vez por parte del sistema,
- Que al conectar directamente a los buses del procesador digital que envía el dato el primer *buffer*, el ruido de estas líneas no se acople al convertidor D/A ya que su dato de entrada presente está en el segundo.

23.3.3.2 Convertidores múltiples

Los convertidores D/A se fabrican también en forma múltiple en un solo "chip" lo que para muchas aplicaciones es muy útil al ahorrar espacio en las tarjetas de circuitos impresos, hacer más bajo el coste por unidad y menos complejo el sistema de interfaces. Por ejemplo, en forma de circuitos integrados podemos encontrar convertidores múltiples de 4, 8, 16 CD/A para resoluciones de 8 y 12 bits, con tiempos de establecimiento de muy pocos microsegundos hasta el novedoso AD5379, de Analog Devices, de 40 canales monótonos, 14 bits, bipolar, entrada serie (hasta 33 MHz de frecuencia de reloj) y paralelo con actualización simultánea de la salida de los 40 canales, salidas de tensión bipolares con 40 amplificadores.

23.3.4 COMPARACIÓN Y CRITERIOS DE SELECCIÓN

Hemos realizado una incursión por los diferentes tipos de convertidores D/A y apuntado en cada caso sus características más relevantes. De ellas se desprende que uno de los convertidores más rápidos para altas resoluciones y con muy buena linealidad, inherentemente monótono, es el convertidor Sigma-Delta. Sin embargo, para aplicaciones en las que no sea importante tan alta resolución, los convertidores de redes R-2R y sus combinaciones con arquitecturas segmentadas ofrecen muy buenas soluciones. Para resoluciones entre 8 y 12 bits, en las que la linealidad sea un elemento de gran importancia, los convertidores inherentemente monótonos exhiben las mejores características.

Todos los convertidores incorporan en sus circuitos interruptores, lógica y amplificadores y estos pueden ser fabricados con diferentes tecnologías. Resulta conveniente que tengamos presente que la tecnología bipolar construye amplificadores, fuentes de corriente y referencias de muy buenas características, pero para los interruptores y la lógica no es la mejor solución y el consumo es generalmente elevado. Por su parte los construidos con CMOS poseen bajos consumos y excelentes características en los interruptores, condensadores y la lógica, pero los amplificadores y las referencias no están a la altura de los bipolares. Es por ello que los procesos que combinan ambas tecnologías (BiCMOS, LC^{2}MOS) aunque con un precio y complejidad del circuito algo mayor, logran un buen balance de estas características.

Seleccionar un convertidor D/A para una aplicación no es una tarea sencilla. Los parámetros más importantes a considerar en la elección son:

- **Número de bits y código de su entrada:** Se debe determinar el número mínimo de bits que garanticen la resolución requerida para la señal analógica. Esto definirá en primera instancia la resolución del convertidor (8, 12, 16 bits, etc.). Si la señal analógica de salida debe ser bipolar, entonces el con-

vertidor deberá responder a un código binario bipolar, siendo el binario con *offset* uno de los más generalizados. Aunque si la salida es de corrientes complementarias, podrá utilizarse un CD/A de código binario natural seguido de un amplificador de instrumentación, como se muestra en la figura 23.18. Para los códigos de mitad de escala (01...11 y 10...00) ambas corrientes serán muy parecidas, diferenciándose en 1 LSB por lo que las salidas del amplificador de instrumentación serán las correspondientes al "cero negativo" y al "cero positivo", respectivamente y para los extremos de escala (11...11 y 00...00) el amplificador de instrumentación ofrecerá las salidas a fondo de escala positiva y negativa, respectivamente.

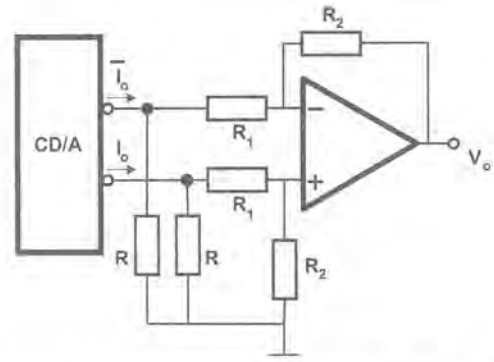


Figura 23.18. Salida bipolar a partir de un código unipolar.

- **Frecuencia de conversión:** el diseño del sistema de medición incorpora ciertas exigencias sobre la frecuencia del refresco del dato al convertidor de donde se obtiene la frecuencia de conversión que el convertidor debe cumplir. Por ejemplo, si se trata de reproducir una señal senoidal de f = 20 Hz mediante 100 muestras por ciclo, la frecuencia de refresco del dato de entrada al CD/A será de 2 kHz y se requiere un CD/A con una frecuencia de conversión mayor de 2 kS/s.
- **Forma en que el dato será transferido:** según las características del procesador que entrega el dato este puede ser enviado al convertidor en serie o en paralelo. El convertidor debe presentar compatibilidad para la transferencia de los datos sin necesidad de ningún hardware adicional.
- **Alimentación disponible:** aunque este es un aspecto que puede ser solucionado, en las aplicaciones con sistemas alimentados con baterías y equipos portátiles pudiera ser de gran importancia. Los fabricantes ofertan convertidores de tecnologías de muy bajo consumo y que se alimentan con una sola fuente de baja tensión, hasta convertidores que necesitan dos fuentes de polaridades opuestas. Aunque en la mayoría de los diseños se debe procurar que el convertidor consuma la menor energía posible.

Existen otras características tales como el tipo de salida analógica (tensión o corriente) y su margen estático de valores que, como dependen del sistema al que se le entrega la señal, generalmente no son satisfechas directamente por el convertidor y requieren un circuito de interfaz y/o la utilización de una fuente de referencia V_{ref} externa para el convertidor. El diseñador tiene que hacer finalmente un análisis comparativo entre las posibles soluciones y elegir la mejor, según un criterio técnico-económico.

23.4 Conversión Analógica-Digital

La conversión analógica-digital es un procedimiento a través del cual una señal analógica de tensión o de corriente unipolar o bipolar es transformada en un código digital N obtenido a partir del cociente entre la

magnitud de la señal analógica de entrada y la de una referencia de tensión o de corriente. El código N correspondiente a una señal de entrada V_i puede obtenerse por:

$$N = \frac{V_i}{V_{ref}} = \sum_{i=1}^{\infty} \frac{b_i}{2^i} \text{ , donde } b_i \in [0,1] \tag{23.22}$$

Este cociente se realiza en binario y el resultado exacto de la conversión se lograría luego de un número infinito de bits.

En los sistemas reales que realizan este procedimiento, llamados **convertidores analógico-digitales** (CA/D), se limita el número de bits a **n** por lo que se realiza la **cuantificación** de la señal analógica en 2^n niveles. Estos niveles se expresan en un código binario, por lo que el convertidor A/D realiza la **codificación** de la señal.

23.4.1 CARACTERÍSTICAS DE LA CONVERSIÓN A/D

23.4.1.1 Características estáticas

La figura 23.19 presenta la característica de transferencia estática de un sistema que realiza la conversión analógica-digital de una señal unipolar, su error de cuantificación y el esquema simplificado de un convertidor A/D con entrada y fuente de referencia de tensión, las señales digitales de control inicio de conversión (*Start Conversión*, SC) y fin de conversión (*End Of Conversión*, EOC).

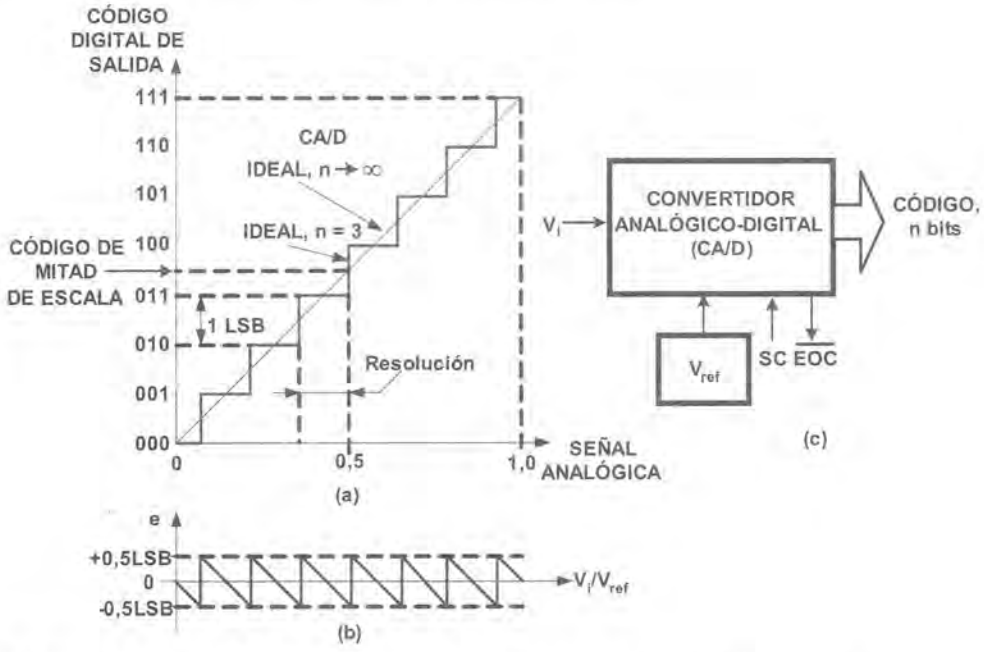


Figura 23.19. a) Característica de transferencia estática de un sistema de conversión analógica-digital ideal, b) error de cuantificación del convertidor ideal de 3 bits, c) convertidor analógico-digital.

La línea continua en a) representa la función de transferencia correspondiente a un convertidor de infinitos bits, mientras que la escalonada representa la de un convertidor (cuantificador) ideal uniforme (véase apartado 23.1) que ofrece la salida en un código unipolar binario de 3 bits (8 códigos) y la entrada se presen-

ta normalizada respecto de su valor máximo (que en los convertidores reales de tensión suele hacerse coincidir con V_{ref}). Puede observarse que:

- Mientras que para el convertidor de 3 bits existe un margen de valores de la entrada para los que la salida se mantiene en un mismo código (cuya magnitud es el peso en términos de la entrada de 1 LSB), en el convertidor de infinitos bits existe un código distinto para cada posible valor de la entrada;
- Para la tensión correspondiente a la media escala existe igual probabilidad de que la salida sea 011 ó 100 en el convertidor de 3 bits;
- El inevitable error de cuantificación (figura 23.19b) es uniforme, de amplitud 0,5 LSB, valor máximo dado por la expresión $e_{máx} = \pm V_{ref}/2^{n+1}$. En la medida que n aumenta, estos parámetros caracterizadores del error de cuantificación se reducen; para n tendiendo a infinito, la amplitud del error tiende a cero.

Los parámetros fundamentales que caracterizan estáticamente a un convertidor A/D son su **resolución**, **margen de entrada**, **ganancia** y **exactitud**.

Resolución (*resolution*): variación necesaria en la tensión de entrada para provocar el cambio entre códigos adyacentes en el código de la salida. Para un convertidor uniforme, si N es el número total de códigos de la salida y FS el fondo de escala, la resolución se cuantifica por:

$$\text{Resolución} = \frac{FS}{N-1}$$

y si N es muy grande puede aproximarse a

$$\text{Resolución} \approx \frac{FS}{N} \tag{23.23}$$

Ilustremos esto con un par de ejemplos:

- Un convertidor A/D de salida en código unipolar binario de 12 bits y FS = 10 V tiene una resolución de $10\ V/(2^{10}-1) = 9{,}7752$ mV, o de $10\ V/2^{10} = 9{,}7656$ mV en forma aproximada.
- Un convertidor A/D de salida en código NBCD 4 décadas y FS = 10 V tiene una resolución de $10V/(10^4-1) = 1{,}0001$ mV, o de 1 mV en forma aproximada.

Muchos fabricantes de convertidores A/D identifican la **resolución** con el **número de bits** del convertidor.

Margen de entrada (*input range*): se refiere al conjunto de valores de la entrada para los que el convertidor ofrece sus códigos de salida. Por ejemplo: 0 ÷ 10 V (unipolar); ±5 V (bipolar).

Ganancia (*gain*): es la pendiente de la recta que une los puntos medios de los peldaños primero y último de la característica transferencial estática del convertidor. En el convertidor ideal la ganancia es la unidad.

Exactitud absoluta (*absolute accuracy*): la exactitud absoluta o error total de un convertidor A/D se refiere a la diferencia entre el código que se obtiene en la salida para una determinada tensión de entrada y el código esperado. Se expresa en LSB o en tanto por ciento del valor de plena escala de la entrada e incluye todos los errores estáticos del convertidor. Por ejemplo, en un convertidor de 16 bits la exactitud absoluta es ±4 LSB equivalente a ±0,006104%. Los fabricantes suelen especificar el valor máximo de este parámetro.

Exactitud relativa (*relative accuracy*): es la diferencia entre el código que se obtiene en la salida para una determinada tensión de entrada y el código esperado luego de quitar los errores de cero y de ganancia. Por ejemplo, en un convertidor de 16 bits la exactitud relativa es ±3 LSB o ±0,004578%. Cuando se expresa el valor máximo de este parámetro, coincide con la no-linealidad integral (INL).

EJERCICIO 23.9

Un convertidor de 16 bits, con entrada de tensión en los márgenes 0 ÷10 V, 0 ÷ 5 V y 0 ÷2,5 V y salida en binario directo posee la tabla siguiente de asociación entre tensiones características y códigos en hexadecimal.

a) Calcule el valor de las tensiones de entrada para cada caso.

b) ¿Qué diferencia existe entre su característica estática y la ideal para 16 bits?

c) Calcule la ganancia.

Tensión de entrada	Código
Fondo de escala – 2 LSB	FFFF
Media escala + 1 LSB	8000
Media escala	7FFF
Media escala – 1 LSB	7FFE
Cero + 2 LSB	0001
Cero	0000

SOLUCIÓN:

a) Si el convertidor es de 16 bits y se comporta como un cuantificador uniforme, el peso de 1 LSB o resolución será:

Para el margen de entrada $0 \div 10$ V 1 LSB $= 10/2^{16} = 152{,}6\ \mu V$
$0 \div 5$ V 1 LSB $= 5/2^{16} = 76{,}3\ \mu V$
$0 \div 2{,}5$ V 1 LSB $= 2{,}5/2^{16} = 38{,}15\ \mu V$

Los valores de tensión en voltios correspondientes a los códigos y la condición de la entrada en cada margen serán:

Condición	Código	V_i, (0 ÷ 10V)	V_i, (0 ÷ 5V)	V_i, (0 ÷ 2,5V)
Fondo de escala – 2 LSB	FFFF	9,999695	4,999847	2.499924
Media escala + 1 LSB	8000	5,000153	2,570076	1,257038
Media escala	7FFF	5,000000	2,500000	1,250000
Media escala – 1 LSB	7FFE	4,999847	2,499924	1,249962
Cero + 2 LSB	0001	0,000305	0,000153	0,000076
Cero	0000	0,000000	0,000000	0,000000

b) Hay tres diferencias importantes:

En primer lugar se ha asignado un código (7 FFF_H) para la tensión correspondiente a la mitad de la escala. En un convertidor ideal no existe un código para la mitad de la escala en unipolar binario (figura 23.19); existe igual probabilidad de que la salida se encuentre en alguno de los dos códigos adyacentes a la mitad de la escala de códigos de salida (en este caso 7 FFF_H y 8000_H) para la tensión de entrada correspondiente a la mitad de la escala. Esto es una práctica muy habitual en la utilización de convertidores A/D porque, en muchos circuitos prácticos, existe la posibilidad de cierto ajuste de su característica estática.

En segundo lugar, con una tensión de 2 LSB menor que la de fondo de escala ya se tiene el código de mayor peso, que debe corresponder a la tensión de fondo de escala y arribar a este con una tensión superior a la de fondo de escala menos ½ LSB.

En tercer lugar, para pasar del código 0000_H al 0001_H hay que aumentar la entrada desde 0 V hasta 2 LSB, mientras que en el convertidor ideal este cambio de código debe suceder cuando la entrada sea mayor que ½ LSB.

c) Si suponemos que el centro del peldaño correspondiente al código de mayor peso se encuentra en el fondo de escala de la entrada menos 1 LSB y que el centro del peldaño del cero en 1 LSB, se tiene que los $2^{16}-1$ espacios entre niveles de cuantificación de la salida se corresponden sólo con $2^{16}-(1+2)$ fragmentos de tensión de valor 1 LSB de la entrada. La ganancia del convertidor será:

$$G = (2^{16}-1)/(2^{16}-3) = 1{,}0000305.$$

La simetría de la característica estática siempre es recomendable para que el convertidor presente iguales errores del cero y a fondo de escala y la ganancia sea la unidad.

23.4.1.2 Características dinámicas

Un convertidor A/D no ofrece de forma inmediata el código de salida correspondiente a la señal aplicada a su entrada. El convertidor inicia el proceso de conversión (figura 23.19c) cuando se le activa su entrada digital de control **SC** y, luego de transcurrida una cierta duración, el convertidor activa su salida digital de control **EOC** informando que ya es estable el dato resultado de la conversión en su salida. Esta duración puede estar comprendida entre pocos nanosegundos y algunos segundos, en dependencia del tipo de convertidor. Los parámetros fundamentales que caracterizan el comportamiento dinámico de un convertidor A/D son el **tiempo de conversión,** la **frecuencia de conversión** y el **ancho de banda de plena potencia** que analizaremos a continuación.

Tiempo de conversión (*conversion time*) T_C: es el tiempo necesario para que el convertidor realice una conversión. En algunas técnicas de conversión en las que se produce un proceso de adquisición y retención del valor de la señal previo a la conversión propiamente dicha, incluye esta duración. Es el tiempo que transcurre entre el inicio de conversión y el fin de conversión.

Frecuencia de conversión (*throughput rate*): es la máxima frecuencia a la cual el convertidor puede ofrecer datos estables en su salida en el modo de trabajo de conversiones consecutivas[5]. En general, es menor o igual al inverso del tiempo de conversión y se expresa en hertzios o en muestras por segundo (S/s). Por ejemplo, el CA/D AD7663 de 16 bits tiene una frecuencia de conversión de 250 kS/s y su $T_C = 4$ µs. A menudo se suele también denominar frecuencia de muestreo (*sampling rate*) del convertidor ya que determina la máxima frecuencia a la que el convertidor puede "muestrear" una señal.

EJERCICIO 23.10

Calcule la mínima frecuencia de muestreo del convertidor necesaria para realizar la conversión A/D de una señal de tensión senoidal de frecuencia f_i = 100 Hz y de amplitud A = 5 V con un convertidor de 10 bits, para que el cambio de la señal de entrada entre dos muestras consecutivas no exceda a 1 LSB.

[5] En este modo de trabajo el convertidor utiliza su señal de salida EOC para crear la de entrada SC, de manera tal que terminando una conversión inicia automáticamente la siguiente. Como durante el tiempo de conversión permanece en su salida el dato resultado de la conversión anterior, sólo impide el acceso a este registro durante su refresco.

SOLUCIÓN:

La mayor rapidez de cambio de tensión en este tipo de entrada se produce en el cruce por cero de la senoide (donde la derivada es máxima) y será:

$$(dV_i/dt)_{máx} = 5 \cdot 2\pi \cdot 100 \text{ V/s}$$

La resolución aproximada del convertidor será: $10 \text{ V}/2^{10} = 0{,}009765 \text{ V} = 1 \text{ LSB}$. Por lo que la frecuencia mínima de muestreo que hace que el cambio entre muestras consecutivas sea de 1 LSB será: $f_{smín} = (5 \cdot 2\pi \cdot 100 \text{ V/s})/ 0{,}009765 \text{ V} = 3{,}2172 \cdot 10^5 \text{ Hz} = 321{,}72 \text{ kHz}$.

La frecuencia de conversión del convertidor debe ser mayor que 321,72 kHz. Obsérvese que esta frecuencia de muestreo es más de 1500 veces mayor que la frecuencia de muestreo del Teorema de Nyquist.

Ancho de banda de plena potencia (*full-power bandwidth*) FPBW: se corresponde con aquella frecuencia de la señal de entrada para la cual la amplitud de la fundamental de la señal reconstruida a su salida se reduce en 3 dB teniendo la entrada amplitud constante e igual al fondo de escala. Por ejemplo, el CA/D AD1672 tiene un FPBW = 5 MHz. Limita el trabajo del convertidor a altas frecuencias de la señal de entrada por lo que en alguna literatura se analiza como un error dinámico.

23.4.1.3 Errores en los convertidores A/D

En los convertidores A/D se presentan un conjunto de errores estáticos y dinámicos dependientes de las componentes de sus circuitos, sus derivas temporales y térmicas, así como de la estabilidad de los circuitos de reloj internos que ofrecen la base de tiempo necesaria y sincronizan su funcionamiento.

Entre los errores estáticos fundamentales pueden citarse el **error de no-monotoneidad**, el **error de cero**, el **error de ganancia**, y los **errores de linealidad integral y diferencial** [23.12].

El **error de no-monotoneidad** tiene la misma interpretación que en los convertidores D/A. Se presenta cuando al incrementar en más de 1 LSB la tensión de entrada el código de salida no pasa a ser el adyacente superior al presente o cuando pasa a ser el adyacente inferior.

La figura 23.20 presenta las características estáticas de convertidores reales exhibiendo exageradamente **errores de cero** (*offset*) en a) y de **ganancia** en b).

El **error de cero unipolar** es el desplazamiento horizontal de la característica transferencial estática y puede cuantificarse calculando la desviación del valor de tensión de entrada V_i para el que se produce el cambio del código 00...00 al código 00...01 del valor correspondiente a ½ LSB, según la expresión:

$$E_{cero} = \frac{V_{i00..01}}{V_{iLSB}} - \frac{1}{2} \text{LSB} \qquad (23.24)$$

El error de cero bipolar es la diferencia entre la tensión de entrada ideal para obtener el código correspondiente al cero en trabajo bipolar y la tensión que realmente produce ese código y se expresa en LSB.

El **error de ganancia** es la diferencia entre los puntos de ganancia[6] nominal y real del convertidor luego de haber ajustado el error de cero. Se expresa en términos de LSB y puede calcularse de forma aproximada mediante la expresión:

---

[6] El punto de ganancia es el valor de tensión de entrada para el que el convertidor de resolución 0 ($n \rightarrow \infty$) ofrece la combinación de salida a fondo de escala.

$$E_{\text{ganancia}} = \frac{V_{iGN} - V_{iGR}}{V_{iLSB}} \tag{23.25}$$

donde V_{iGN}, V_{iGR} son las tensiones de entrada correspondientes a los puntos de ganancia nominal y real del convertidor.

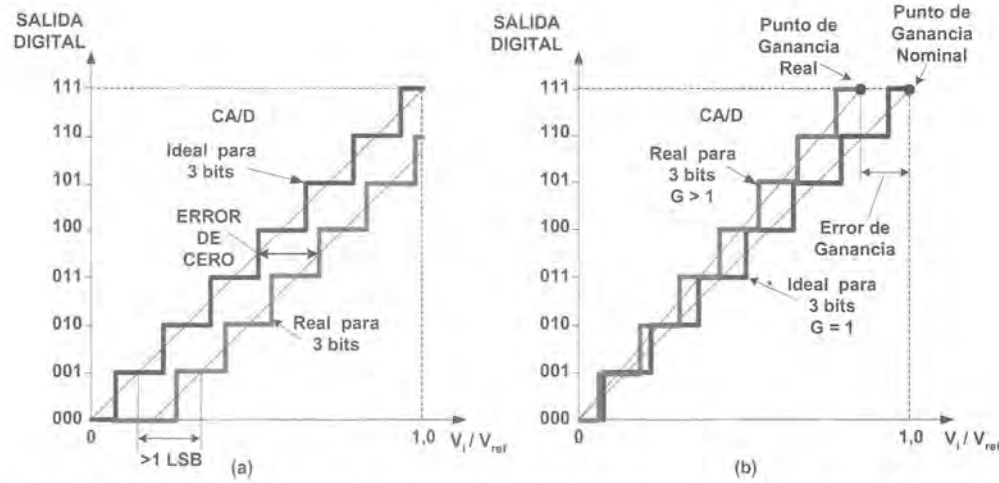


Figura 23.20. Ilustración de los errores de cero unipolar y de ganancia en un CA/D de 3 bits.

En muchos convertidores resulta posible ajustar los errores de cero y de plena escala, con lo que se ajusta la ganancia.

La figura 23.21 presenta exageradamente los errores de linealidad de un convertidor de 3 bits y la forma en que influyen en el error de cuantificación del convertidor.

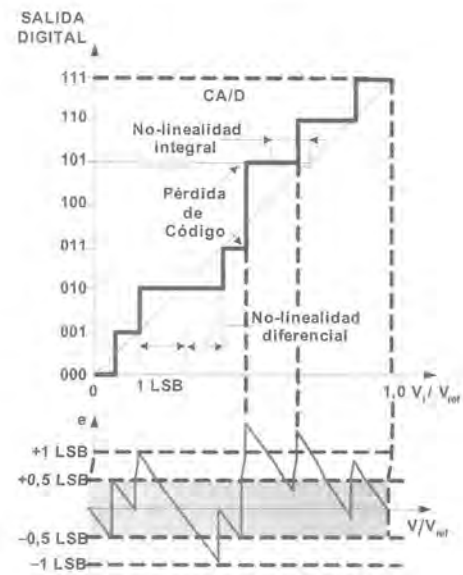


Figura 23.21. Ilustración de los errores de linealidad y el error de cuantificación para un CA/D de 3 bits.

El **error de no-linealidad diferencial** (DNLE) es la diferencia entre el ancho del peldaño de la característica estática real y su ancho ideal de 1 LSB. Se expresa en términos de LSB o en tanto por ciento del margen de la entrada. Ejemplo: Un convertidor de 16 bits con un margen de entrada $V_{iFSR} = 10$ V tiene una no-linealidad diferencial máxima $DNL_{máx} = \pm 0{,}25$ LSB.

Cuando el error de no-linealidad diferencial es mayor que 1 LSB el convertidor presentará no-monotoneidad. Incluso puede presentarse el caso ilustrado en la figura 23.21 de **pérdida de códigos** (*missing codes*).

El **error de no-linealidad integral** (INLE) es la máxima desviación de la característica estática del convertidor de la línea recta que une los puntos de cero y de ganancia luego de ajustar a cero los errores de cero y de ganancia. A menudo se refiere este error simplemente como **linealidad**, y se suele expresar en LSB o en tanto por ciento del margen de entrada. Por ejemplo, para un convertidor de 16 bits, $V_{iFSR} = 10$ V, INLE = ±3 LSB o 0,004578% del margen de entrada.

EJERCICIO 23.11

Un convertidor A/D de 12 bits con salida en unipolar binario para un margen de entrada 0 ÷ 5 V presenta un error de no-linealidad integral de ±1 LSB. Para una entrada $V_i = 2{,}35$ V, ¿qué código se obtiene a la salida?

SOLUCIÓN:

El peso en voltios de 1 LSB será: $1\ LSB = 5\ V/2^{12} = 0{,}001221$ V.

Si el convertidor fuera ideal, el código correspondiente a una entrada de 2,35 V en decimal será:

$N_d = 2{,}35/0{,}001221 = 1925{,}12 = 1925 + 0{,}12$. Como la fracción no llega a 0,5, entonces el código en binario será el correspondiente a 1925. (En el caso de ser la fracción igual a 0,5, existe igual probabilidad de que la salida se encuentre en el código binario correspondiente a 1925 ó 1926 y, si es mayor, el código sería el correspondiente a 1926.)

$$N = 011110000101 \rightarrow 785_H$$

Como el convertidor tiene un error de no-linealidad integral de ±1 LSB, entonces la salida podrá ser 784_H, 785_H o 786_H.

Las derivas térmicas de los errores anteriormente citados también constituyen una fuente de error cuando el convertidor trabaja en un amplio margen de temperatura. Su interpretación es similar a la analizada en los convertidores D/A.

Entre los errores dinámicos del convertidor producidos por la no-idealidad de la característica transferencial dinámica del sistema podemos citar, además del inherente al proceso de cuantificación ya estudiado, el **error de apertura**, la **distorsión armónica total** producida, la **relación señal-ruido+distorsión** y, como consecuencia de esta última, el **número efectivo de bits** a los que haremos referencia a continuación.

Error de apertura: este error es producido por la incertidumbre en el instante en el que el sistema de muestreo-retención (S/H) que entrega su señal al convertidor pasa del modo muestreo al de retención, como se presenta en la figura 23.22.

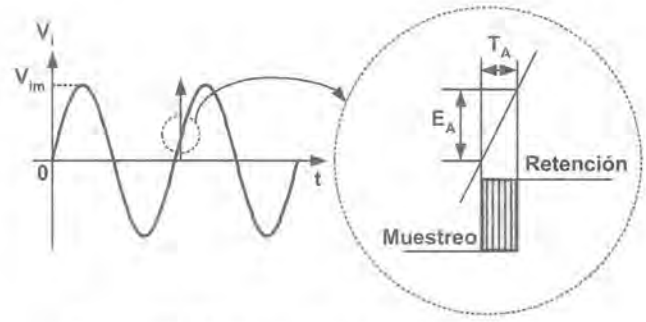


Figura 23.22. Ilustración del error de apertura.

Esta variación T_A se debe al ruido en el reloj del sistema (*sampling clock jitter*) y/o al ruido en la señal y durante ese tiempo se produce una modificación del valor de la señal de entrada de valor E_A. Si esta variación se "sale" del marco de ½ LSB en el punto de máxima rapidez de cambio de la señal, afectará la exactitud del convertidor. De esta forma, este error impone una limitación a la máxima frecuencia de la señal de entrada que puede calcularse conociendo la incertidumbre del tiempo de apertura del S/H del sistema.

Para una señal senoidal de entrada como la mostrada en la figura 23.22, la máxima rapidez de cambio ocurre en su cruce por cero, donde será:

$$\left|\frac{dV_i}{dt}\right|_{máx} = 2 \cdot \pi \cdot f_i \cdot V_{im}$$

Para que el error de apertura no afecte a la exactitud del convertidor, debe ser menor que ½ LSB en el punto de máxima rapidez de cambio de la entrada y si el convertidor tiene n bits, se tendrá que:

$$E_A = T_A \cdot \left|\frac{dV_i}{dt}\right|_{máx} \leq \frac{1}{2} LSB = \frac{2 \cdot V_{im}}{2^{n+1}}$$

y de ambas expresiones se puede obtener, para la condición "igual a", el valor de la frecuencia máxima de la señal de entrada hasta la cual no se produce error de apertura en el convertidor mediante:

$$f_{imáx} = \frac{1}{T_A \cdot \pi \cdot 2^{n+1}} \tag{23.26}$$

EJERCICIO 23.12

Si el valor de la incertidumbre del tiempo de apertura del S/H de un sistema con un convertidor de 16 bits es de ±0,5 ns, ¿cuál es la frecuencia máxima de la señal de entrada para la cual el error de apertura aún no afecta a la exactitud de la conversión?

SOLUCIÓN:

El valor de la incertidumbre del tiempo de apertura es: $T_A = 10^{-9}$ s.

De la ecuación (23.26) se tiene:

$$f_{imáx} = \frac{1}{10^{-9} \cdot \pi \cdot 2^{17}} = 2{,}428 \text{ kHz}$$

A partir de esta frecuencia, el sistema S/H+convertidor disminuirá su relación señal-ruido producto del aumento de este último por este error. La calidad binaria real disminuirá al reducirse el número efectivo de bits y, con ello, la exactitud del convertidor.

Los convertidores A/D monolíticos actuales reducen este efecto al incorporar en el circuito el sistema de muestreo y retención eliminando los problemas de ruido del reloj. Por ejemplo, el CA/D AD7663 de 16 bits posee un retardo de apertura[7] de 2 ns, pero su incertidumbre (la que provoca el error de apertura) es de sólo 5 ps(rms). Con esta incertidumbre el convertidor podrá trabajar con señales de entrada de hasta 17,17 MHz sin afectar su exactitud por este error.

Distorsión armónica total (*total harmonic distortion*) THD: es el cociente entre la suma efectiva de los valores rms de los primeros cinco armónicos que tiene la señal reconstruida a la salida y el valor rms de la tensión de plena escala de la entrada. Como depende de las no linealidades del convertidor, este parámetro depende de la frecuencia de la señal de entrada, y se expresa en dB. Por ejemplo, el CA/D AD1672, tiene generalmente $THD_{f=100kHz} = -74$ dB y $THD_{f=1,5MHz} = -60$ dB.

Relación señal-ruido+distorsión (*signal-to-noise+distortion*) S/N+D: es la relación señal-ruido del convertidor (ver ecuación 23.9) incluyendo el ruido provocado por la distorsión del convertidor de frecuencias menores a la de Nyquist debida a la no-linealidad de su característica estática y a otros efectos. Se expresa en dB y depende de la frecuencia de la señal de entrada f_i. Ejemplo: El CA/D AD7663, cuyas hojas de características se encuentran en el CD, posee una *S/N+D = 90 dB para* f_i *= 10 kHz, mientras* que para f_i = 100 kHz es de sólo 30 dB.

Número efectivo de bits (*effective number of bits*) ENOB: este parámetro define la calidad binaria real del convertidor. Para un convertidor ideal de n bits con entrada senoidal de amplitud igual al fondo de escala, la relación señal-ruido será (ver apartado 23.1):

S/N(dB) = 10log6 + 20nlog2 = 7,782 + (6,021)n, [dB]. Si tenemos un CA/D con una S/N de X dB, entonces el número efectivo de bits del convertidor será:

$$\text{ENOB} = \frac{X - 1,76}{6,02} \tag{23.27}$$

EJERCICIO 23.13

Calcule el número efectivo de bits del CA/D AD7663 de 16 bits para a) f_i *= 10 kHz, b)* f_i *= 100 kHz, considerando sólo el valor de su relación señal-ruido + distorsión.*

SOLUCIÓN:

Aplicando la expresión (23.27) se obtiene:

a) $\text{ENOB}_{fi=10\ kHz} = (90 - 1,76)/6,02 = 14,65$ bits, si se aproxima al entero menor: 14 bits.
b) $\text{ENOB}_{fi=100\ kHz} = (30 - 1,76)/6,02 = 4,7$ bits, aproximado será: 4 bits.

De los resultados podemos observar que, en la medida que aumenta la frecuencia de la señal de entrada, disminuye el número efectivo de bits producto de la no-idealidad del convertidor.

[7] El retardo de apertura sólo introduce un retardo de igual magnitud entre el momento que se toma la muestra de la señal y el momento en que realmente comienza la conversión. Este parámetro debe tenerse muy en cuenta cuando se usan varios CA/D sincronizados.

Existen otros errores de menor trascendencia tales como la distorsión de intermodulación y el margen dinámico libre de señales espurias que sugerimos al lector los consulte en la literatura especializada de convertidores A/D o, de forma más escueta, en las propias hojas de datos de los fabricantes.

23.4.2 TIPOS DE CONVERTIDORES

Atendiendo al procedimiento por el cual, a partir de la tensión de entrada, se obtiene el código de salida, los convertidores A/D pueden clasificarse en dos grandes grupos: **directos** e **indirectos.** Los **directos** son los que obtienen directamente el código de salida a partir de la tensión de entrada, mientras que los **indirectos** convierten la tensión de entrada en otra variable (duración de un pulso, frecuencia de un tren de pulsos) y esta es convertida finalmente en un código.

23.4.2.1 Convertidores directos

Entre los convertidores directos se encuentran el convertidor **paralelo** o ***flash***, el **rastreador** y el de **aproximaciones sucesivas.**

Convertidor paralelo (*flash ADC*). La figura 23.23 presenta el esquema básico de este convertidor para obtener un código unipolar binario de n bits en la salida.

La señal de entrada V_{in} es aplicada a los N-1 comparadores ($N = 2^n$) donde es comparada con las tensiones ofrecidas por el divisor de Kelvin modificado, en el que las resistencias de los extremos V_{ref} y masa son de la mitad del valor del resto, para que los peldaños de la característica transferencial estática de cero y plena escala tengan ½ LSB de ancho. Aquellos comparadores en los que V_{in} sea mayor que la tensión del nodo del divisor aplicada a su entrada ofrecerán un estado "1" en su salida por lo que generarán un código de termómetro (ver apartado 23.3.2.2) a la entrada del codificador $2^n - 1 \rightarrow n$ que lo convierte en un código binario de n bits. Este código es almacenado en un registro de salida en el momento en que la señal de control S (*strobe*) —que en su flanco de subida habilita las salidas de los comparadores y al codificador— baje a cero indicando el fin de la conversión (EOC). Estos convertidores son muy rápidos con tiempos de conversión del orden de pocos nanosegundos por lo que inherentemente son convertidores muestreadores pero:

- Requieren 2^n resistencias y 2^n–1 comparadores que, como deben ser muy rápidos, consumen mucha energía.
- Las resistencias del divisor deben ser de pequeño valor para poder suministrar las corrientes de polarización de los comparadores con lo que la fuente de referencia debe suministrar corrientes relativamente altas.
- Poseen una gran capacidad de entrada lo que limita su trabajo a altas frecuencias.

En la práctica, estos convertidores están disponibles como circuitos integrados de 8 bits, con frecuencias de conversión del orden de 500 MS/s y anchos de banda de potencia de hasta 300 MHz. Para menores resoluciones se utilizan de n = 3 ó 4 bits como componente de otras arquitecturas que analizaremos a continuación.

Convertidor de sub-margen o semi-paralelo (*half-flash ADC*). Para lograr resoluciones correspondientes a códigos de 12 bits o inferiores se utiliza en la práctica la arquitectura de sub-margen presentada en la figura 23.24. La señal de entrada es aplicada a un primer convertidor *flash* de p bits cuya salida es convertida en una tensión por el CD/A de p bits de muy alta calidad. Esta tensión difiere de la de entrada en una cantidad comprendida entre 0 y 1 LSB del primer convertidor y es restada de la señal de entrada.

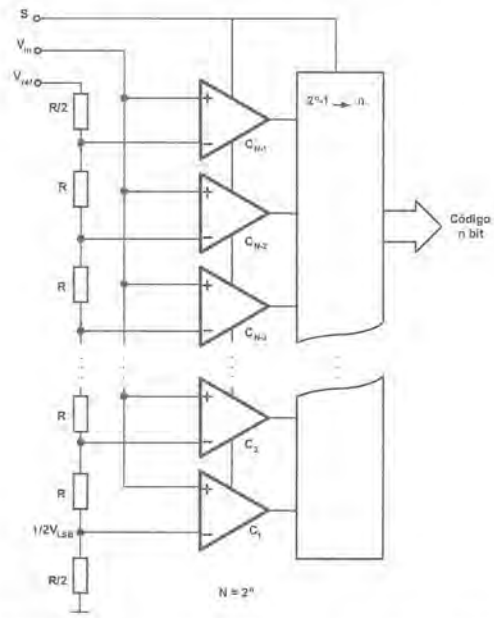


Figura 23.23. Esquema básico del convertidor A/D paralelo.

El resultado es amplificado con una ganancia[8] G y tomado como señal de entrada de un segundo convertidor *flash* de q bits, o del propio primer convertidor. El código de salida del primer convertidor constituye los p bits más significativos del código de salida, mientras que los del segundo los q bits menos significativos.

Es posible realizar una corrección de los errores de la conversión mediante el solapamiento del bit menos significativo del primer convertidor con el más significativo del segundo y tomar la decisión de su valor mediante un sumador completo [23.8] por lo que para obtener el número de bits total n = p + q se necesita que los convertidores sean de p + 1 y q + 1 bits, respectivamente.

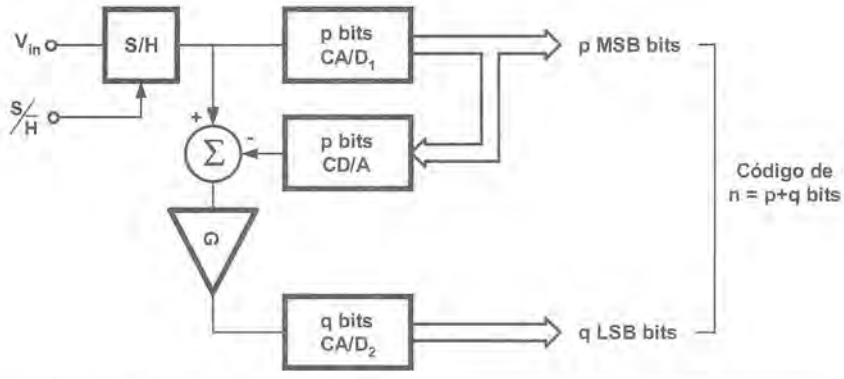


Figura 23.24. Arquitectura básica del convertidor A/D de sub-margen.

[8] $G = 2^p - 1$ si la tensión de referencia del segundo convertidor es la misma que la del primero. En general $G = V_{ref2}/V_{LSB1}$.

Esta estructura puede ser repetida con 4 ó 6 elementos, pero presenta limitaciones derivadas de los atrasos introducidos por el tiempo de establecimiento de los CD/As y la estabilidad de la ganancia de los amplificadores. No obstante, ofrece tiempos de conversión de cientos de nanosegundos para una resolución correspondiente a 12 bits. La figura 23.25 ofrece la arquitectura interna del CA/D AD1671, 12 bits, T_C = 800 ns (Analog Devices) de 570 mW de consumo, que puede ofrecer la salida en código en complemento a dos o en binario con *offset* en forma paralelo.

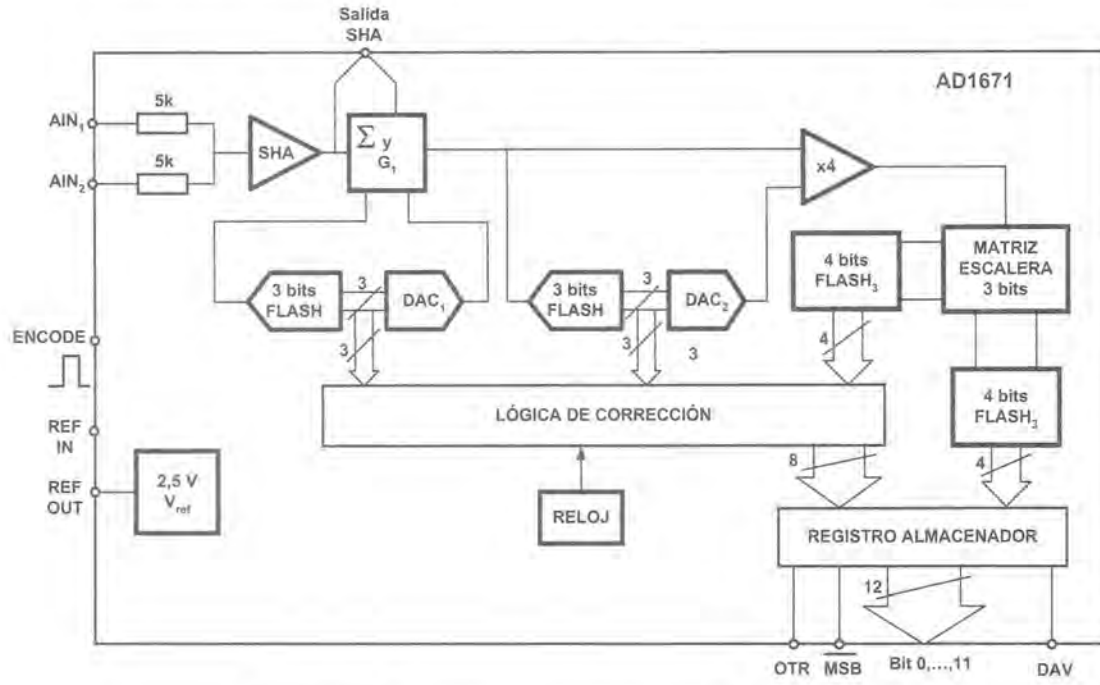


Figura 23.25. Arquitectura del convertidor A/D AD1671.

La conversión se inicia con el flanco de subida de un pulso aplicado a la entrada ENCODE y un reloj interno se encarga del control preciso del S/H, los convertidores flash de 3 y 4 bits y los CD/As. La lógica de corrección obtiene los 4 bits más significativos del resultado a partir de los 6 bits de los dos convertidores flash de 3 bits mediante solapamiento y, los 8 bits menos significativos, de los dos convertidores de 4 bits. Las hojas de características de este convertidor se encuentran en el CD.

Convertidor multi-etapas "*pipelined*". Una modificación de la arquitectura anterior que mejora considerablemente la respuesta dinámica se presenta en la figura 23.26, correspondiente al CA/D AD1672, 12 bits, T_C = 330 ns, construido con tecnología BiCMOS, de sólo 240 mW de consumo, de salida paralelo en código binario directo.

Se basa en incorporar entre las etapas convertidor $flash_n$-CD/A_n un sistema de "rastreo" y retención (THA) (*track-and-hold amplifier*) con ganancia mayor que la unidad.

Al iniciarse la conversión, el THA_1 captura el valor de la señal de entrada presente pasando del modo muestreo al de retención y lo aplica a la primera etapa $flash_1$-CD/A_1. Al propio tiempo, el THA_2 inicia su trabajo en el modo muestreo y en su entrada se produce la resta de la señal de entrada con el resultado de la conversión de la primera etapa. Cuando esta es estable, el THA_2 captura este resultado y, amplificado convenientemente, es llevado a la entrada de la segunda etapa convertidor $flash_2$-CD/A_2, al propio tiempo que el

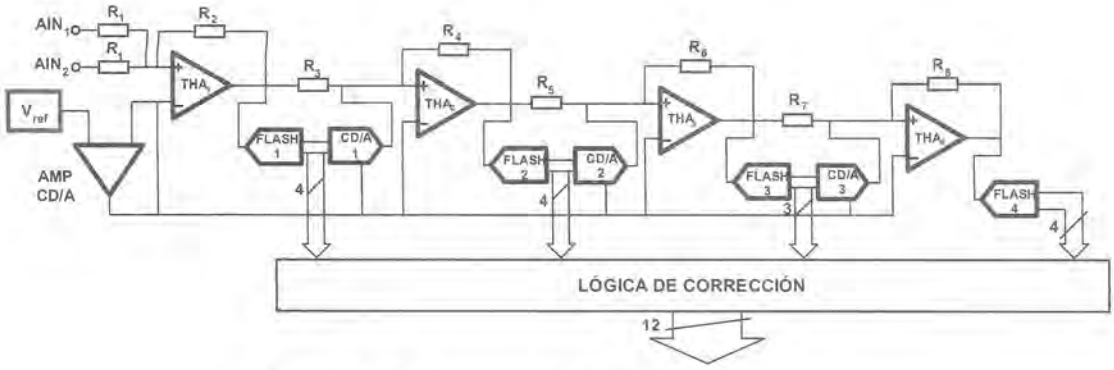


Figura 23.26. Arquitectura multi-etapas *"pipelined"*.

THA_3 inicia su trabajo en el modo muestreo. Este proceso se repite de manera tal que se consigue realizar la conversión en un menor tiempo. La lógica de corrección elabora los 8 bits más significativos a partir de los resultados de las etapas 1, 2 y 3, y los 4 bits menos significativos son el resultado de la conversión del convertidor $flash_4$.

Convertidor rastreador (*tracking ADC*). Si los convertidores flash son rápidos y de tiempo de conversión constante, independientemente de la magnitud de la señal de entrada y de sus cambios entre conversiones consecutivas, el convertidor rastreador es muy rápido y lento a la vez. El tiempo de conversión es variable y depende directamente de la diferencia entre los valores de la señal entre conversiones consecutivas. La figura 23.27a presenta el esquema básico de este convertidor que con sólo un comparador, un CD/A y un contador de pulsos "*up-down*" logra precisiones limitadas por la resolución del CD/A. Siempre que la señal de entrada V_i sea diferente del equivalente en tensión del código existente en el contador V_o, los pulsos del reloj aumentarán o disminuirán el valor del código según el signo de la diferencia sea positivo o negativo, respectivamente. Con el inicio de conversión (figura 23.27b) el contador se limpia y al ritmo del reloj de periodo T cambia su contenido hasta "atrapar" el valor de la señal de entrada y luego "sigue su rastro" en un entorno dependiente de la resolución del sistema. Este convertidor tiene la ventaja de ser muy sencillo, de bajo coste para muy pequeñas resoluciones y de ofrecer permanentemente el resultado de la conversión en el modo de conversiones consecutivas con gran precisión, siempre que la variación de la señal de entrada durante el periodo T del reloj del sistema no exceda a 1 LSB.

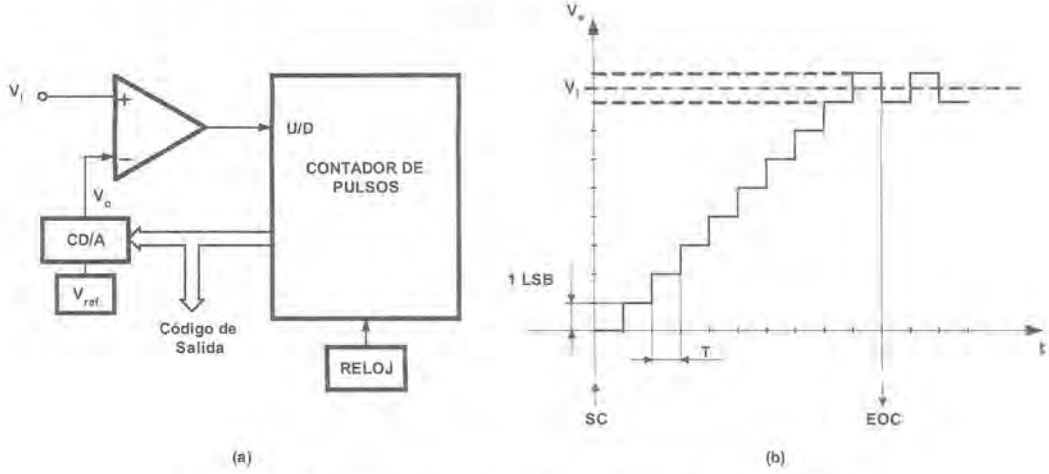


Figura 23.27. Esquema básico del convertidor A/D rastreador.

Sin embargo, presenta las desventajas de un tiempo de conversión variable y muy grande para n superiores a 12 bits y cambios bruscos de gran magnitud en la señal de entrada. Es muy poco utilizado en la práctica, pero con un pequeño cambio en su arquitectura logramos uno de los convertidores de mayor utilización en nuestros días: el de aproximaciones sucesivas.

Convertidor de aproximaciones sucesivas (*succesive approximation ADC*). La figura 23.28 presenta el esquema básico del convertidor de aproximaciones sucesivas. Si se comparan las figuras 23.28a y 23.27a, nos daremos cuenta de que sólo se diferencian en el tipo de sistema que memoriza el código presente.

En este convertidor, al iniciar la conversión un registro de aproximaciones sucesivas (SAR) envía al CD/A el código correspondiente a la mitad de la escala (10000...0) que, en forma de tensión V_o, se corresponde con el peso en voltios del bit más significativo del código. Esta tensión V_{o1} es comparada con la señal de entrada V_i y, si es menor, el bit más significativo del código es "1"; por el contrario si es mayor, entonces será "0". Esta decisión se toma durante el primer ciclo de reloj. Durante el segundo ciclo de reloj se envía por parte del SAR al CD/A el código con el resultado de la decisión anterior en el primer bit y un "1" en el segundo, repitiéndose la toma de decisión en función de la nueva comparación entre V_{o2} y V_i. Siguiendo este algoritmo, la conversión durará tantos ciclos de reloj como bits tenga el código de salida, siendo independiente de la magnitud de la señal de entrada V_i, como nos ilustra la figura 23.28b para un convertidor de 3 bits. Los tiempos de conversión de estos convertidores se encuentran desde cientos de nanosegundos para 8 bits hasta algunos microsegundos para 16 bits y es la arquitectura que más se utiliza en aplicaciones de hasta 12-16 bits de moderada exactitud.

La señal de entrada durante el tiempo de conversión tiene que permanecer **constante**, por lo que en este convertidor es imprescindible el uso de un sistema de muestreo y retención (S/H).

En los circuitos prácticos monolíticos se emplean diversas técnicas para conformar el S/H y el CD/A. Por ejemplo, el CA/D AD7663, 16 bits, T_C = 4 µs, construido con tecnología CMOS con sólo 35 mW de consumo presenta una arquitectura muy peculiar para el convertidor D/A que, a su vez, realiza las funciones de muestreo y retención. El convertidor D/A (figura 23.29) es de capacidades ponderadas de 16 bits que, con los interruptores de rama S_n en la posición representada en la figura, se cargan a una tensión igual a la de la señal de entrada funcionando de esta forma el conjunto de condensadores como un sistema de muestreo y retención en el modo muestreo.

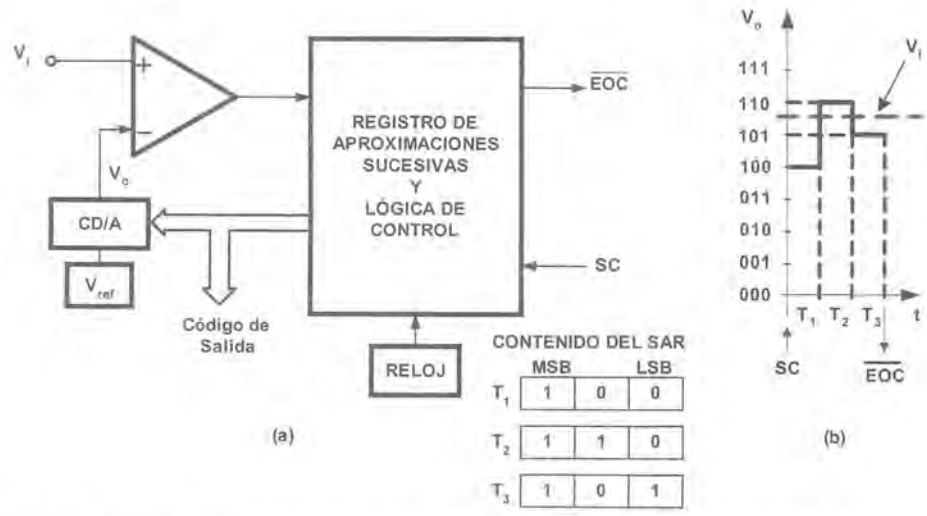


Figura 23.28. Esquema básico del convertidor A/D de aproximaciones sucesivas para n = 3 bits.

El condensador asociado a la entrada inversora del comparador se carga a la tensión de la masa de la señal de entrada. Al iniciar la conversión se abren los interruptores $S_{WA,B}$ y los interruptores S_n y S_D conectan los condensadores a la masa de la referencia de V_r. En forma diferencial aparecerá una tensión a la entrada del comparador equivalente a la de la señal de entrada referida a su masa y este tendrá un primer desequilibrio. *Al ritmo del reloj y, comenzando por el MSB hacia el LSB, la lógica de control actúa uno a uno sobre* los interruptores S_n conectando el condensador de rama a V_{ref} o a masa de la referencia en función de la salida del comparador con lo que la entrada del comparador variará en pasos de $V_{ref}/2$, $V_{ref}/4$, ... $V_{ref}/2^{16}$. Al completar los n ciclos del reloj, en la lógica de control se encuentra el código correspondiente a la entrada. La gran diferencia entre los valores de los condensadores hace que la no-linealidad integral de esta arquitectura sea notable: ±3 LSB en este caso, pero el convertidor es rápido y de muy bajo consumo.

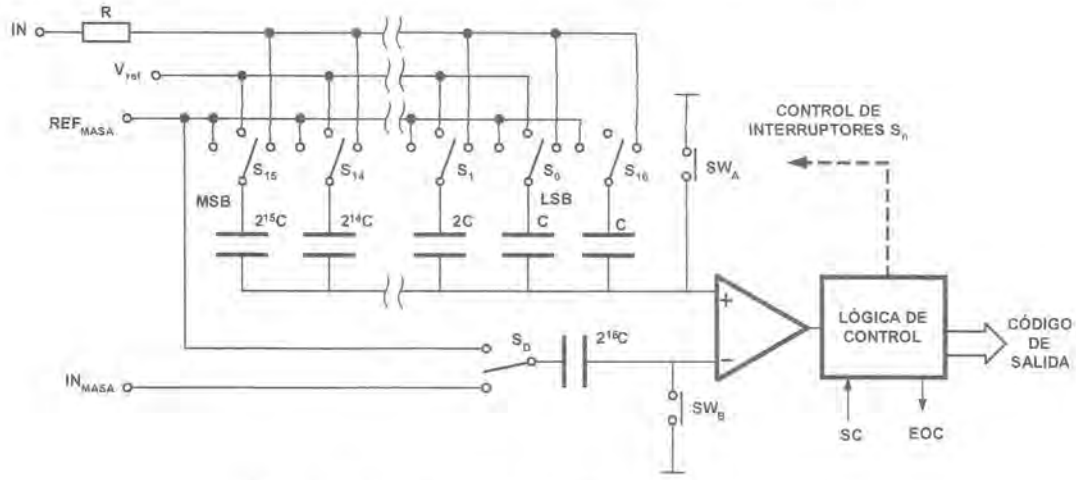


Figura 23.29. Arquitectura interna del CA/D AD7663.

Con este principio de conversión se fabrica una gran diversidad de convertidores que incluyen un multiplexor analógico para 4, 8 y 16 canales simples o diferenciales y sistemas de muestreo-retención y que se denominan como Sistemas de Adquisición de Datos monolíticos, muy útiles para aplicaciones en arquitectura de estrella para el tratamiento de señales de bajas frecuencias. En estos sistemas, es importante tener en cuenta que el tiempo de conversión de la señal correspondiente a un canal incorpora los tiempos de conmutación del multiplexor, de muestreo del SHA, el retardo de apertura y el tiempo de conversión propiamente dicho del convertidor. De esta forma, la frecuencia de muestreo máxima del sistema de adquisición para n canales es equivalente a la de un canal dividida por n.

23.4.2.2 Convertidores indirectos

Entre los convertidores indirectos se encuentran los **integradores** o **de rampas** y los de **sobre-muestreo.** Los integradores se caracterizan por ser lentos pero de muy buena precisión, bajo coste y gran sencillez circuital, siendo los más difundidos los de conversión **tensión-tiempo (V/T)** y **tensión-frecuencia (V/f).** Los de **sobre-muestreo** utilizan como técnica básica la modulación delta y son los convertidores A/D más utilizados hoy día para las aplicaciones de señales de voz (telefonía móvil) y puentes de sensores para señales de gran espectro de frecuencias gracias a la combinación que logran de muy pequeñas resoluciones (20-24 bits) con frecuencias de muestreo de cientos de megahertzios a un bajo coste.

Convertidores de tensión-tiempo (CV/T). Estos convertidores se basan en la medición del tiempo que tarda una rampa de tensión en alcanzar un nivel de tensión dado. Si durante ese tiempo se cuentan pulsos de

un reloj de alta estabilidad en un contador, su contenido será proporcional a la magnitud de la señal de entrada. Las variantes más difundidas en la práctica son las siguientes:

1) **Convertidor de simple pendiente.** La figura 23.30a ilustra el principio de simple integración o simple pendiente: un impulso de inicio de conversión SC "limpia" el contador binario de n bits CT y da inicio a la rampa positiva V_o de pendiente V_{ref}/RC del generador de rampa GR. El comparador habilita la puerta Y que deja pasar hacia el contador los pulsos de reloj de alta frecuencia f_c. En el instante (figura 23.30b) en que la magnitud de la rampa alcance el valor de la tensión de la señal de entrada V_i, el comparador bloqueará la puerta Y y el contenido Z del contador (número de pulsos de frecuencia f_c contados) es directamente proporcional a la magnitud de V_i, pudiendo relacionarse por las siguientes expresiones:

$$\frac{V_i}{T_c} = \frac{V_{ref}}{R \cdot C}$$

y además $Z = T_c \cdot f_c$, de donde finalmente se obtiene:

$$Z = \frac{f_c}{V_{ref} \cdot R \cdot C} \cdot V_i \qquad (23.28)$$

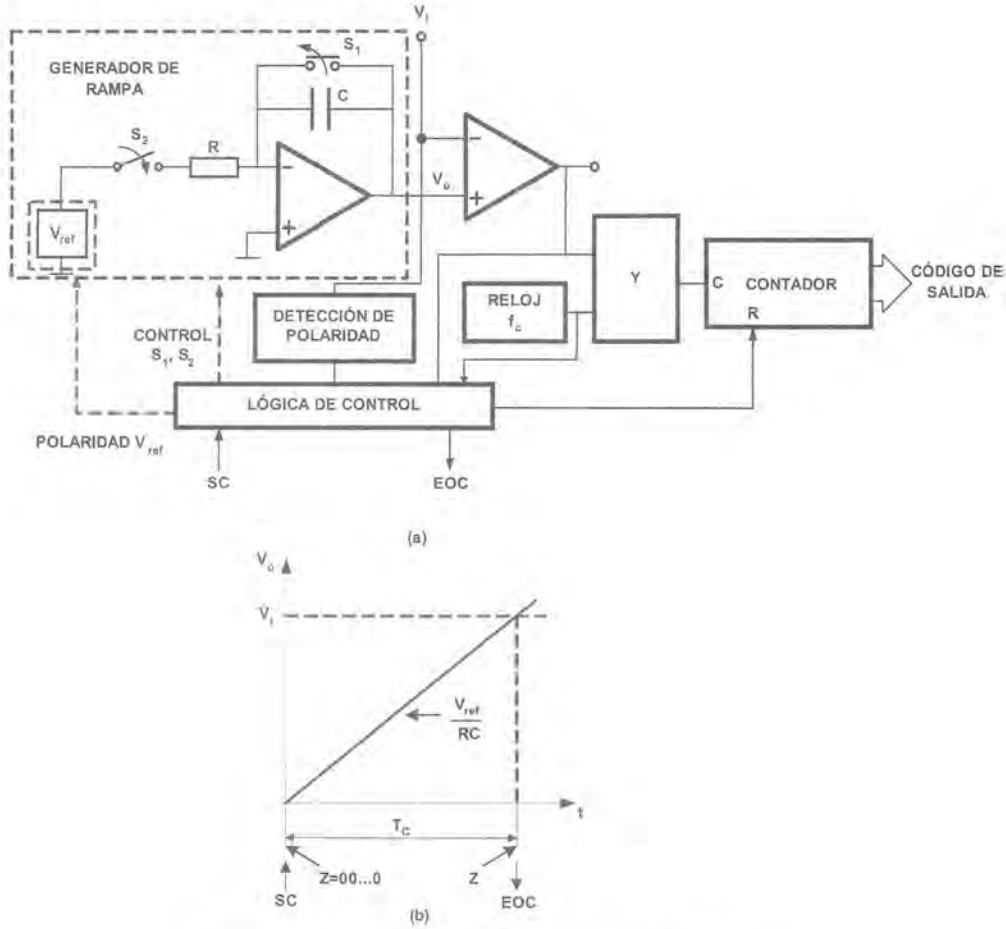


Figura 23.30. Convertidor V/T de simple pendiente.

El tiempo de conversión es dependiente de la magnitud de la señal; su valor máximo ocurrirá para $V_i = V_{ref}$ y podrá evaluarse por la expresión $T_{cmáx} = RC$.

En los convertidores integradores, la polaridad de la tensión de referencia y la señal de entrada tienen que ser opuestas por lo que la lógica de control se encarga de garantizarlo mediante un circuito de detección de la polaridad. En este convertidor se requiere gran linealidad en el generador de rampa y alta estabilidad temporal y térmica en la frecuencia del reloj y en los valores de la resistencia R y el condensador C del integrador. Su inmunidad al ruido que acompaña a la señal de entrada no es alta y el resultado de la conversión es equivalente al valor instantáneo de la señal de entrada en el instante $t = T_c$. Si en ese instante se produce una variación brusca de la señal provocada por el ruido, el resultado de la conversión podrá ser incorrecto. Este convertidor no supera exactitudes del orden de 0,1%.

2) **Convertidor de doble pendiente.** La figura 23.31a presenta la arquitectura simplificada del muy difundido convertidor de doble rampa. Con el inicio de conversión SC, el contador binario se limpia, el interruptor S_1 se abre y S_2 conecta la señal de entrada V_i al integrador. La lógica de control habilita la puerta Y comenzando a pasar los pulsos del reloj al contador. Durante un tiempo fijo T_1 (marcado por el desbordamiento del contador en $N = 2^n$ pulsos) se produce la rampa de pendiente $-V_i/RC$ alcanzando la señal V_o (figura 23.31b) un nivel V_{o1} proporcional al **valor promedio** $\overline{V_i}$ durante T_1 de la señal de entrada V_i.

Al concluir T_1 el contenido del contador es 0...00, la lógica de control conecta la tensión de referencia de polaridad opuesta a la de entrada al integrador y se inicia la segunda integración, ahora con pendiente de valor constante (V_{ref}/RC). Al cruzar por cero V_o el comparador cambia de estado, la lógica de control bloquea la puerta Y y el contenido del contador Z será $Z = T_2 \cdot f_c$.

Pero $f_c = 2^n/T_1$ por lo que la cantidad de pulsos contados durante T_2 será:

$$Z = \frac{T_2}{T_1} \cdot 2^n \tag{23.29}$$

además, en el sistema se cumple que:

$$\frac{\overline{V_i}}{R \cdot C} \cdot T_1 = \frac{V_o}{R \cdot C} \cdot T_2$$

de donde se tiene que:

$$T_2 = T_1 \cdot \frac{\overline{V_i}}{V_{ref}} \tag{23.30}$$

Sustituyendo (23.30) en (23.29) se tiene finalmente:

$$Z = \frac{2^n}{V_{ref}} \cdot \overline{V_i} \tag{23.31}$$

El resultado de la conversión es **independiente** de la frecuencia del reloj y de la constante de tiempo del integrador RC y sólo se requiere estabilidad en sus valores en el marco del ciclo de conversión, lo que resulta técnicamente posible aún con componentes de no muy alta calidad, aunque el tiempo de conversión es del orden de milisegundos o incluso segundos para resoluciones de 16 bits.

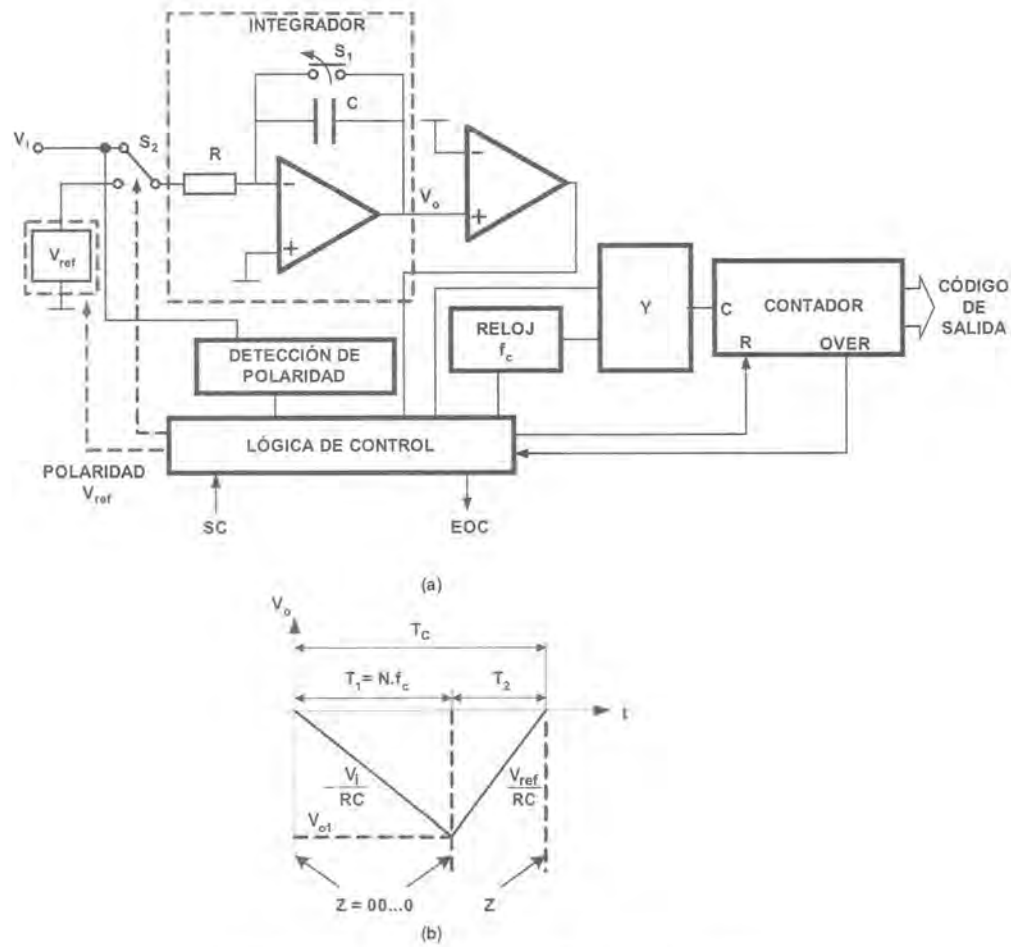


Figura 23.31. Convertidor V/T de doble pendiente.

Particular importancia tiene el hecho de que la señal de entrada es promediada durante la primera integración para su aplicación en voltímetros digitales. Si T_1 se hace coincidir con un múltiplo del periodo de la señal de ruido que acompaña a la tensión a convertir (rizado debido a la tensión de la red eléctrica de 50 Hz o 60 Hz), su efecto es teóricamente eliminado. La figura 23.32 presenta la característica de atenuación del ruido acompañante a la señal en función de su frecuencia. Existen polos de atenuación en cada múltiplo de $1/T_1$ y, en general, existe una característica de atenuación importante para otras frecuencias de ruido.

3) **Convertidor de triple pendiente.** La figura 23.33a ilustra la técnica de integración de triple pendiente con la que se consiguen las mismas ventajas que en la de doble pendiente con un aumento relativo de la complejidad del circuito, pero se logra reducir el tiempo de conversión al decidir los bits más significativos del código de salida con una mayor pendiente.

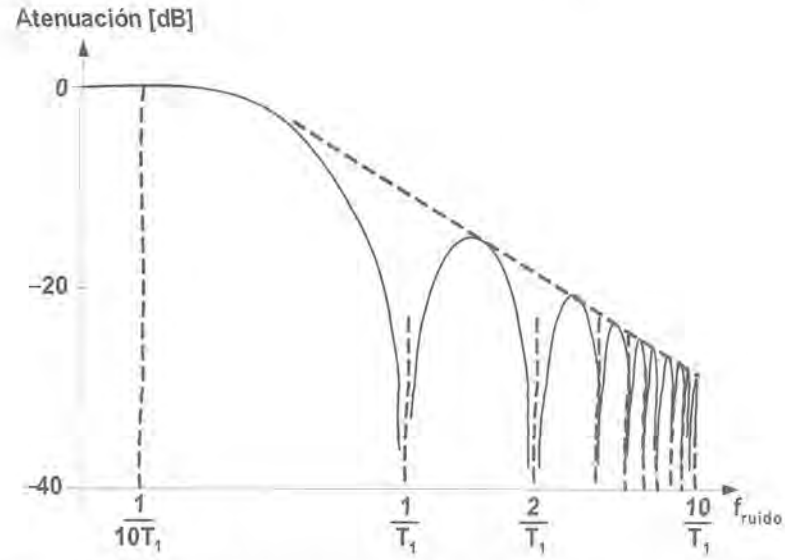


Figura 23.32. Característica de atenuación del ruido del convertidor A/D que integra la señal durante T_1.

Se utilizan dos valores de tensión de referencia $V_{ref1} > V_{ref2}$ y, en el comparador C_1, una tensión de comparación $-V_H$ igual al peso en voltios del máximo contenido del contador LSB. El proceso de integración de la señal V_i se realiza durante T_1 temporizado por el desbordamiento del contador MSB + LSB, instante en el que la entrada se conmuta a $-Vref_1$ si y sólo si la magnitud $V_o > V_H$. Si esta condición no se cumple, entonces la entrada se conmutará a $-V_{ref2}$ (figura 23.33b para $V_i = V_{i2}$) y el contenido del contador MSB se mantiene en 0...00. Si suponemos que se cumple (por ejemplo, $V_i = V_{i1}$), la integración de $-V_{ref1}$ durará hasta que V_o alcance el nivel $-V_H$ dirigiéndose los pulsos del reloj hacia el contador MSB durante T_2 y el contenido de MSB será el correspondiente a los bits más significativos del código de salida. En este momento ($T_1 + T_2$ en la figura 23.33b) comienza la integración de $-V_{ref2}$ y los pulsos del reloj se dirigen al contador LSB cuyo contenido en el cruce por 0 de V_o se corresponderá con los bits menos significativos del código. Con este principio se construyen convertidores de múltiples rampas que logran muy altas resoluciones (14-18) bits con tiempos de conversión del orden de cientos de milisegundos a segundos. Existen versiones en las que, en lugar de utilizar dos fuentes de referencia, se utilizan dos valores de resistencias R para provocar el cambio de las pendientes.

Convertidores de tensión-frecuencia (V/f). Estos convertidores, a menudo conocidos como osciladores controlados por tensión (VCO), se basan en realizar integraciones consecutivas de la señal de entrada partiendo desde cero hasta alcanzar un nivel de referencia con una muy pequeña constante de tiempo RC, requiriendo muy buena linealidad en el integrador por lo que, en algunos de los tipos existentes, la señal de entrada es convertida en una corriente para cargar el condensador a corriente constante. Son convertidores de bajo *coste con altas resoluciones (14 bits).*

Entre los tipos más difundidos se encuentran el de **multivibrador controlado por corriente** y el de **equilibrio de carga**, a los que nos referiremos a continuación.

1) **Convertidor V/f de multivibrador controlado por corriente.** Realmente este convertidor es un convertidor de corriente-frecuencia ya que convierte la señal de entrada en una corriente (figura 23.34a).

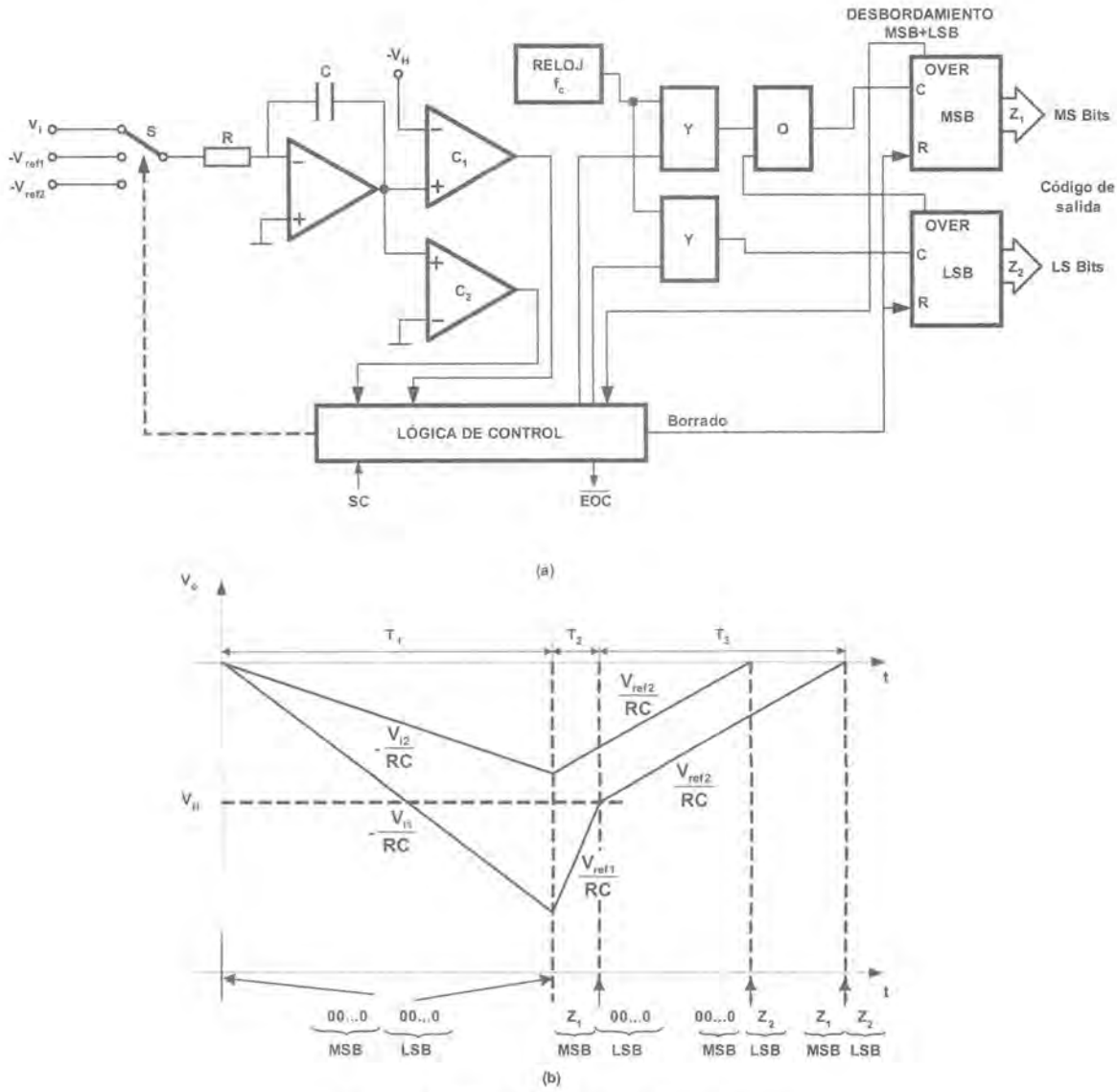
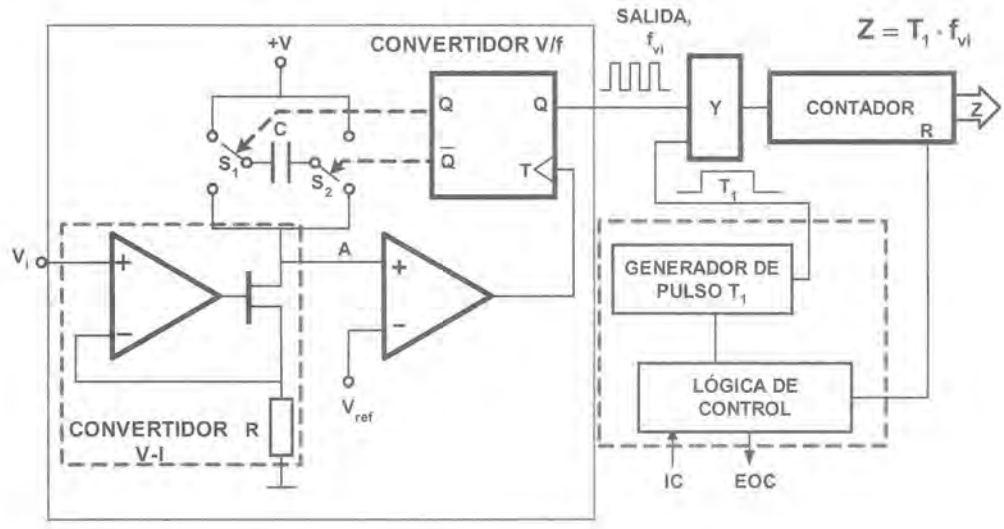


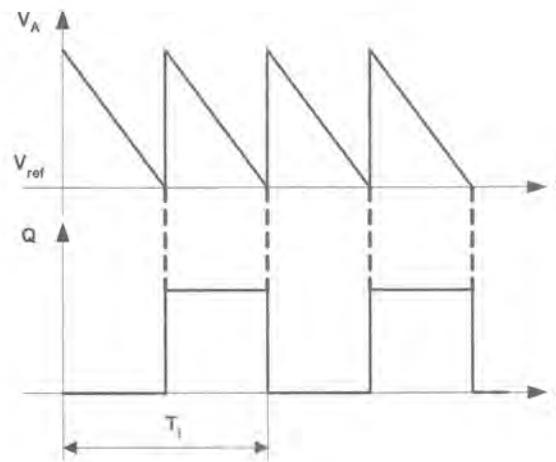
Figura 23.33. Convertidor de triple rampa.

La pendiente de la rampa de V_A depende de la magnitud de la corriente de entrada (figura 23.34b) y, por ende, la frecuencia de la onda cuadrada de la salida del multivibrador será directamente proporcional a V_i. Con el inicio de conversión SC el generador de pulsos T_1 abre la puerta Y dejando pasar al contador el tren de pulsos del multivibrador, cuya frecuencia depende directamente de la tensión V_i. Al finalizar T_1 el contenido Z del contador será $Z = T_1 \cdot f_{Vi}$. Este convertidor puede lograr resoluciones correspondientes a 14 bits y los límites de sus posibilidades son establecidos por el ruido umbral del comparador y el coeficiente térmico de ese umbral, así como la estabilidad dieléctrica del condensador. Su principal desventaja estriba en la falta de sincronismo entre el reloj que produce la duración T_1 y los pulsos que son contados, pero como circuito integrado se oferta a muy bajo coste con gran capacidad para trabajar con diferentes tensiones de alimentación +V y V_{ref} lo que lo hace ideal para aplicaciones de telemetría de media precisión, donde el convertidor V/f

puede encontrarse muy cerca del sensor y muy alejado del contador. También, si los pulsos del convertidor V/f son transferidos a través de un opto-acoplador, se puede obtener un alto aislamiento entre la salida y la entrada del convertidor.



(a)



(b)

Figura 23.34. Convertidor V/f de multivibrador controlado por corriente.

Convertidor V/f de equilibrio de carga. La versión de este convertidor más utilizada en la practica es la síncrona (*synchronous VFC*) a la que haremos referencia a continuación. La figura 23.35 presenta el esquema básico de este convertidor en el que el condensador del integrador se carga a partir de la señal de entrada V_i hasta que alcance el nivel de comparación V_{ref}.

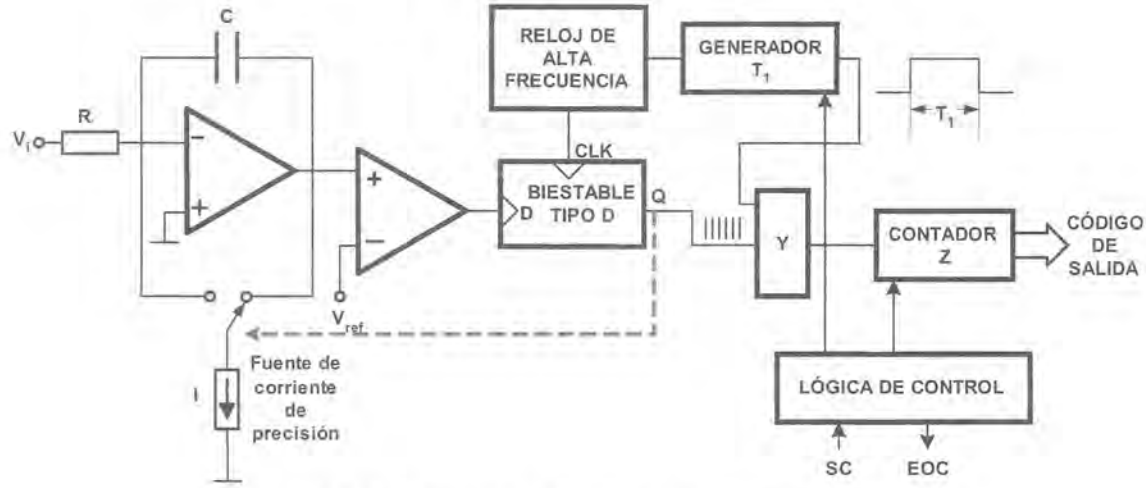


Figura 23.35. Convertidor V/f de equilibrio de carga síncrona.

En este instante, la fuente de corriente de precisión drena una cantidad fija de carga pero la corriente debida a la señal de entrada continúa fluyendo durante la descarga de manera que no se pierde carga alguna de la entrada. La carga fija que aumenta o disminuye el condensador está determinada por la magnitud de la corriente I y la duración del estado del biestable, que siempre será un múltiplo exacto del periodo del reloj de alta frecuencia y síncrono con él. De esta forma, la frecuencia de los pulsos de salida del biestable será exactamente proporcional a la rapidez con la que el condensador del integrador se cargará desde la señal de entrada a su magnitud. La limitación a bajas frecuencias la impone la estabilidad de la fuente de corriente y, a altas frecuencias, la conmutación del interruptor y la frecuencia del reloj.

Cuando un convertidor V/f trabaja en gran señal, su frecuencia de salida es muy alta y, aun cuando se realice el recuento de estos pulsos durante pequeños T_1, se puede lograr una gran resolución pero, para señales pequeñas, resulta muy difícil obtener altas resoluciones. Si el periodo de los pulsos de salida es grande, este puede utilizarse como T_1 y hacer pasar pulsos de reloj al contador (conversión tensión-tiempo) y lograr también altas resoluciones para pequeñas señales. Este tipo de arquitectura combinada la utiliza el VFC AD1170 de resolución desde 7 hasta 18 bits con tiempos de conversión programables desde 1 ms hasta 350 ms y excelente comportamiento a frecuencia de conversión de 50 Hz para lograr un alto rechazo al rizado de la red eléctrica. Las hojas de características de este convertidor se ofrecen en el CD.

Convertidores de sobre-muestreo (*over-sampling ADC*). En síntesis, estos convertidores cuantifican la señal de entrada con la más baja resolución posible (con un simple comparador o convertidor de 1 bit) pero a muy alta frecuencia de muestreo (cientos de MHz). Su frecuencia de muestreo es K veces mayor ($K = 20 \div 512$) que la frecuencia de Nyquist $f_s = 2 \cdot f_m$ (véanse figura 23.36a y capítulo 22) para una señal dada. Esto hace que el ruido de cuantificación se distribuya en un ancho de banda mayor ($Kf_s/2$) y su densidad de potencia se reduzca drásticamente, como se presenta en la figura 23.36b.

Si se aplica este resultado a un filtro digital paso bajo, quitaremos en gran medida el ruido de cuantificación sin afectar el espectro de la señal de entrada, pudiendo reducir entonces la frecuencia del resultado por un factor M (figura 23.36c) con un procedimiento digital de "supresión" (*decimation*) [23.9] cumpliendo aún ampliamente con el criterio de Nyquist, pero con un resultado de un número de bits equivalente muy elevado (ENOB = 18 ÷ 24 bits) al aumentar considerablemente la relación señal-ruido reduciendo este último.

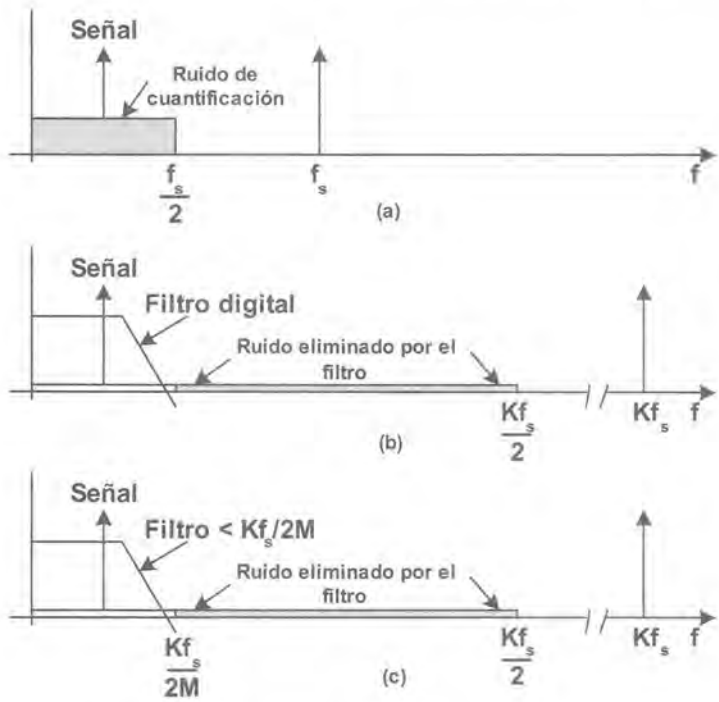


Figura 23.36. Procesos de sobre-muestreo, filtrado digital y decimación.

Los circuitos prácticos de estos convertidores utilizan la modulación delta [23.10] y son conocidos como **convertidores sigma-delta** [23.11]. La figura 23.37 presenta la arquitectura interna de un convertidor sigma-delta de primer orden. La salida A del comparador (convertidor A/D de 1 bit) es la salida del modulador

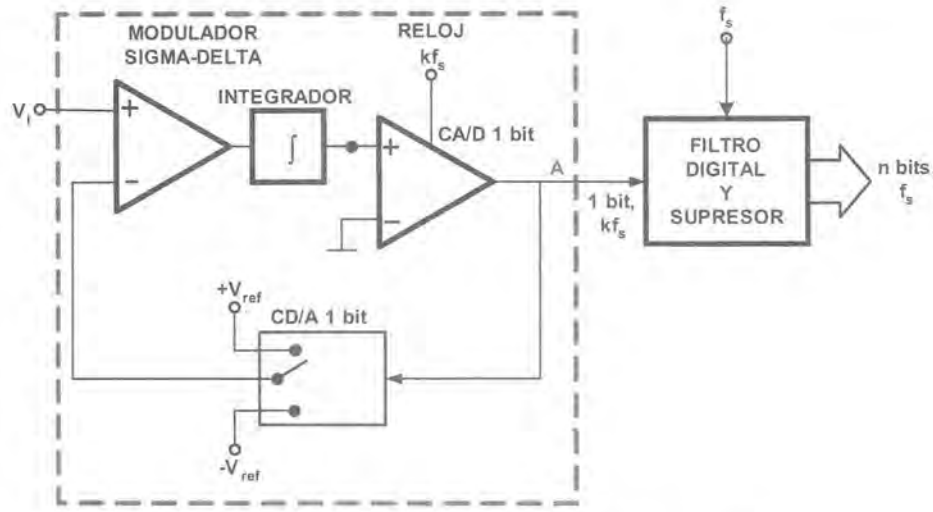


Figura 23.37. Convertidor sigma-delta de primer orden.

en forma de una cadena de pulsos de frecuencia Kf_s que se aplican al convertidor D/A de 1 bit. En dependencia de que el pulso presente a su entrada sea un "0" o un "1", el DAC aplicará a su salida una tensión $+V_{ref}$ o $-V_{ref}$, respectivamente, que se resta del valor actual de la señal V_i en el amplificador diferencial. Este resultado se aplica al integrador que trabaja como en el convertidor V/f de equilibrio de carga, donde se trata de mantener la carga neta del condensador muy próxima a cero, equilibrando las cargas inyectadas por V_i y la tensión de referencia. De esta forma, el promedio local de la señal de salida del modulador "rastrea" con mayor precisión la señal de entrada en la medida que la frecuencia de muestreo Kf_s sea mayor. Si la salida del modulador se aplica a un filtro digital paso bajo y a un supresor (*decimator*), el resultado es la cuantificación del valor de V_i con una muy alta resolución.

El modulador sigma-delta configura el ruido de cuantificación de tal manera que lo ubica por encima del corte del filtro de salida y el número efectivo de bits (ENOB) del convertidor es mucho mayor que lo que se esperaría con sólo aumentar la frecuencia de muestreo K veces. Si se añaden otras etapas de suma y otros integradores en el modulador (figura 23.38a), se consiguen aún mejores resultados en la reducción del ruido de cuantificación y, por ende, mejor relación S/N para una misma frecuencia de muestreo Kfs (figura 23.38b), pero para órdenes superiores a dos, los moduladores ya no ofrecen características lineales [23.14].

Estos convertidores exhiben una gran linealidad, no requieren sistema de muestreo-retención, ofrecen resoluciones de 18-24 bits pero no pueden ser utilizados para procesar varios canales debido a que necesitan la historia pasada de la señal para su funcionamiento. Sin embargo, si se utilizan circuitos de muestreo-retención para cada canal, es posible su utilización siempre que se tenga en cuenta el tiempo de establecimiento de los filtros internos. Un ejemplo de este tipo de convertidor es el AD7710, 24 bits, ±0,0015% de exactitud en 23 bits, amplificador de ganancia programable por software entre 1 y 128 en 8 pasos con entra-

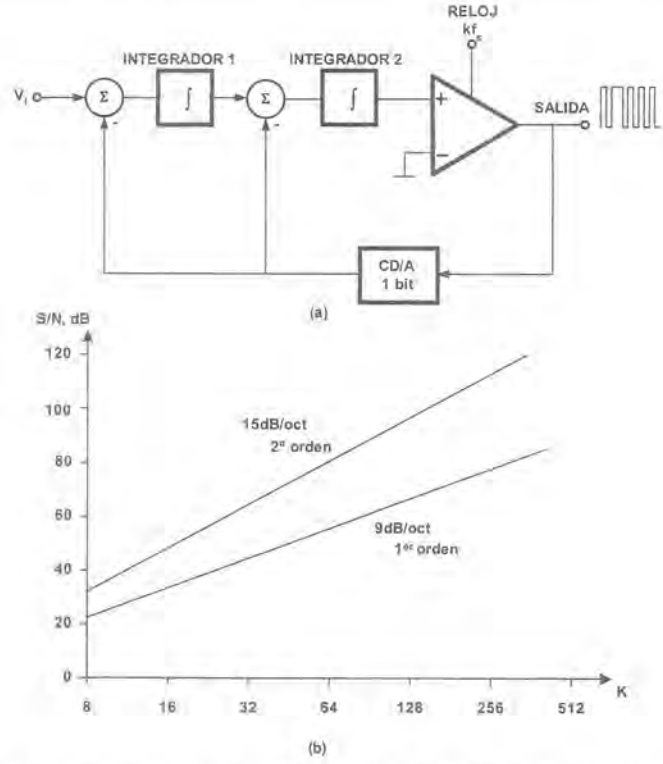


Figura 23.38. a) Convertidor sigma-delta de segundo orden; b) características de S/N en función de K del convertidor.

da diferencial. La frecuencia de corte del filtro digital es programable en 9 valores (de acuerdo con el espectro útil de la señal) y sus frecuencias de conversión pueden alcanzar hasta 312 kHz para una frecuencia del reloj de 10 MHz.

La figura 23.39 presenta un esquema simplificado de este convertidor en una aplicación con un puente de galgas extensométricas alimentado con corriente. La resistencia R_1 se selecciona de manera tal que $R_1 = V_{ref} / I$, siendo I la corriente que excita el puente. El amplificador de ganancia programable hace posible que el margen analógico de entrada de plena escala sea tan bajo como 20 mV y sus entradas diferenciales permiten tensiones de modo común de hasta los niveles de las fuentes V_{DD} y V_{SS}.

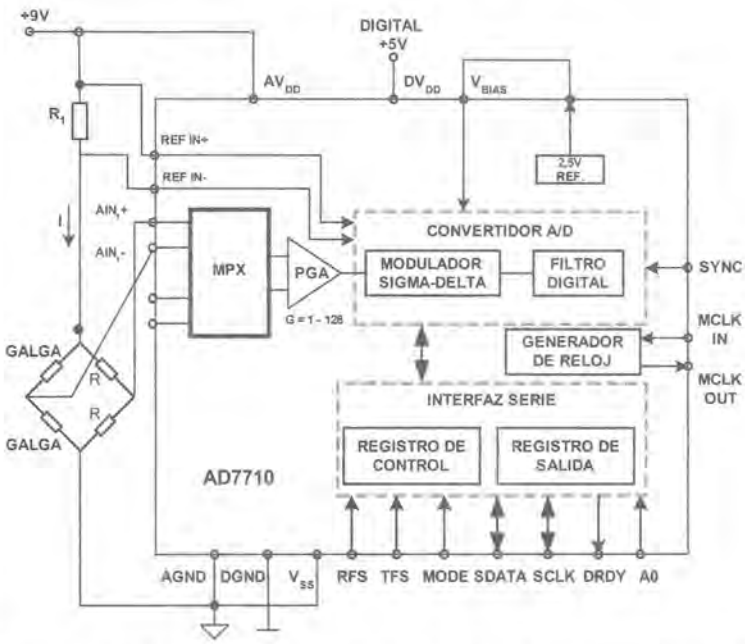


Figura 23.39. Aplicación del AD7710 con un puente de galgas extensométricas. (Analog Devices)

23.4.3 COMPARACIÓN Y CRITERIOS DE SELECCIÓN

Como hemos podido observar en el apartado anterior, el campo de los convertidores A/D es muy amplio, tanto en diversidad de principios de conversión como en técnicas para realizarla. La arquitectura utilizada para implementar el convertidor está caracterizada por el número de bits (vinculado directamente con la resolución) y la frecuencia de conversión; una primera comparación entre los tipos de convertidores se puede obtener al relacionar estos dos parámetros y la arquitectura [23.13]. La figura 23.40 presenta una relación entre el número de bits del convertidor y la frecuencia de muestreo para los tipos más difundidos de convertidores.

Podemos observar que las menores resoluciones (mayor número de bits en el código de salida) se consiguen con los sigma-delta, mientras que las mayores frecuencias de conversión se consiguen con los convertidores paralelo. La inclinación del eje de las elipses quiere dar la idea de que, en un mismo tipo de convertidor y tecnología con la que se fabrica, el logro de una mayor frecuencia de conversión generalmente implica una reducción en el número de bits del convertidor.

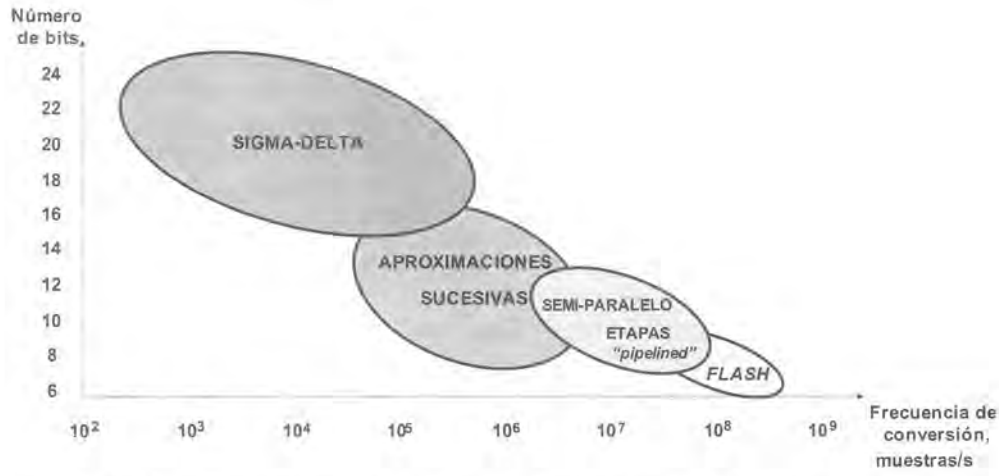


Figura 23.40. Comparación entre los tipos de convertidores de frecuencia de muestreo independiente de la señal.

Los convertidores sigma-delta van ganando cada día más terreno en las aplicaciones de instrumentación cuando se requiere lograr una reproducción digital de señales analógicas de amplio espectro, como el audio, con gran calidad binaria a muy bajo coste.

Los convertidores de aproximaciones sucesivas son los que se utilizan más ampliamente en los sistemas de adquisición de datos ya que ofrecen "resoluciones" entre 8 y 16 bits con frecuencias de conversión moderadamente altas y salida paralelo, incorporando el sistema de muestreo-retención y en muchos circuitos el multiplexor analógico para 4, 8 y 16 canales simples o diferenciales, con costes y consumos de energía relativamente bajos.

Los convertidores de arquitecturas semi-paralelo y los de multietapas "*pipelined*" presentan un buen compromiso entre muy bajo consumo/coste y elevadas frecuencias de conversión, con resoluciones correspondientes a 12-14 bits, mientras que los paralelos o flash, por su gran consumo de energía quedan confinados solamente a aquellas aplicaciones en las que sean imprescindibles frecuencias de conversión muy altas con resoluciones altas (hasta 8 bits).

Los convertidores tensión-tiempo tienen tiempos de conversión variable con la señal de entrada y esta particularidad los relega sólo a aplicaciones de medida de valor promedio de la señal durante tiempos relativamente grandes (decenas de milisegundos hasta varios segundos) y con precisiones altas (correspondientes a 12-18 bits).

Los convertidores tensión-frecuencia tienen un gran campo de aplicaciones en instrumentación, sobre todo en la medida remota de magnitudes físicas lentas, como la temperatura y la presión.

La elección de un tipo de convertidor para una aplicación no es un problema muy complejo si se tienen bien definidas las exigencias de frecuencia de conversión máxima y la resolución que se requiere en la conversión. Estos parámetros, junto con otros factores, tales como el tipo de señal de entrada (unipolar o bipolar), el formato (unipolar o bipolar binario) y el tipo (paralelo o serie) de la salida, la compatibilidad con el procesador o sistema al que va a entregar el dato, la alimentación disponible y el coste, pueden ofrecernos un conjunto de características que darán como resultado una buena elección dentro de la abundante oferta de los fabricantes.

Resumen

- La conversión entre variables analógicas y digitales se realiza mediante circuitos electrónicos especiales denominados convertidores analógico-digitales (CA/D) y digital-analógicos (CD/A) y el resultado de la conversión no es una reproducción fiel de la entrada ya que se necesitaría un número de estados o niveles infinito para la variable digital lo que, en la práctica, no es posible.
- En la conversión D/A existen errores estáticos de no-linealidad que hacen que a la salida aparezcan variaciones temporales y térmicas de la tensión o corriente correspondiente a un código de entrada fijo. También existen limitaciones frecuenciales en el refresco del código de entrada motivadas, en primer lugar, por el tiempo de establecimiento y otros errores dinámicos. Estas variaciones dependen del tipo de circuito que realiza la conversión y de su tecnología de fabricación.
- En la conversión A/D se realiza la cuantificación y codificación de la señal analógica (generalmente de tensión) en 2^n códigos diferentes de salida por lo que existe un error inherente a la cuantificación que hace que un margen de valores de la entrada obtenga un mismo código como salida. Además, los errores relacionados con las no-linealidades temporales y térmicas de la característica estática provocan que, para una tensión fija en la entrada, la salida pueda tener diferentes códigos. Estos errores dependen del tipo de circuito y la tecnología de fabricación del convertidor.
- En los convertidores A/D, el código correspondiente a la entrada presente se produce luego de un transitorio denominado tiempo de conversión que limita su frecuencia de muestreo y que depende del principio de conversión del convertidor. Los de arquitectura paralelo son los más rápidos, mientras que los de rampas tensión-tiempo son los que presentan los mayores tiempos de conversión que depende de la magnitud de la señal.

Problemas propuestos

23.1 Para un cuantificador uniforme de 512 estados calcule: a) la resolución real, b) la resolución aproximada, c) el error de cuantificación máximo, d) el número de bits del cuantificador si los niveles se expresan en un formato binario directo, e) el número de digitos si los niveles se expresan en un formato decimal.

23.2 A un cuantificador uniforme de margen de entrada $0 \div 5$ V y 1024 estados se aplica una señal de continua. a) Calcule la relación señal-ruido (S/N) máxima del cuantificador para este tipo de entrada, b) calcule la S/N que corresponde a una señal de 2,85 V.

23.3 Un convertidor D/A de 10 bits unipolar binario tiene salida en tensión en el margen de $0 \div 10$ V y utiliza una fuente de referencia de 10 V. a) Si para el código de plena escala la salida es de 9,9975 V, ¿cuál es su ganancia? b) Si su exactitud es de ±0,007%, ¿qué tensión de salida se puede esperar para la entrada $0FE_H$?

23.4 El convertidor D/A del problema anterior posee un error de no-linealidad de ±2 LSB a 25°C y un coeficiente térmico de este error de ±1 ppm FSR/°C. Calcule el valor de la salida esperada para la entrada $17F_H$ a 60°C.

23.5 Un convertidor D/A acepta en su entrada códigos binarios con *offset* de 12 bits, el margen de salida es ±5 V y su frecuencia de conversión máxima es de 50 MHz. ¿Cuál es la frecuencia máxima de la

senoide de 5 V de pico que puede ofrecerse a la salida sin que la diferencia en voltios entre dos escalones consecutivos sea mayor de la tensión correspondiente a 1 LSB?

23.6 Un convertidor A/D de margen de entrada 0 ÷ 5 V y salida en código unipolar binario de 10 bits, a) ¿cuántos códigos diferentes puede ofrecer a su salida? b) ¿Qué resolución tiene exactamente? c) ¿Qué resolución aproximada tiene? d) ¿Cuánto es el peso en voltios de 1 LSB? e) ¿Qué código ofrece a su salida para una entrada de 3,235 V si no posee errores de linealidad y la ganancia es 1?

23.7 Si el convertidor A/D del problema anterior tiene una exactitud absoluta de ±3 LSB en todo el margen de temperaturas de trabajo, ¿cuál sería el código de salida esperado?

23.8 A un convertidor A/D ideal de margen de entrada –5 V ÷ 5 V y salida en código complemento a dos de 12 bits se le aplica una senoide de 5 V de pico y frecuencia f = 20 Hz. ¿Cuál es la mínima frecuencia de muestreo del convertidor que garantiza que el cambio de la señal de entrada entre dos conversiones consecutivas no exceda a 1 LSB?

23.9 Un convertidor A/D de margen de entrada 0 ÷ 10 V y salida en código unipolar binario de 16 bits produce el cambio de la combinación 0000_H a la 0001_H al aplicársele a la entrada 2 mV y el cambio de $EFFF_H$ a $FFFF_H$ en 9,9945 V. a) ¿Qué errores de cero y plena escala tiene el convertidor? b) ¿Cuál es su ganancia?

23.10 El convertidor del problema anterior tiene un error de linealidad de ±0,0046% del margen de entrada. Considerando que hemos ajustado los errores de cero y fondo de escala a cero, ¿qué salida se obtendría para una tensión de entrada de 6,875 V?

24

Procesadores digitales de señal

24.1 Microprocesadores, microcontroladores y DSP
24.2 Procesamiento de la información
24.3 Comunicación hombre-máquina
24.4 Sensores inteligentes

24.1 Microprocesadores, microcontroladores y DSP

Microprocesadores (μP)

Un **microprocesador** (μP) es un circuito programable integrado en una única pastilla de silicio que realiza las funciones encomendadas a un **procesador** digital. La figura 24.1 muestra el diagrama de bloques de un procesador digital genérico, donde se distinguen dos partes fundamentales: la **unidad de control**, que interpreta las instrucciones y genera la secuencia de señales para su ejecución, y la **unidad aritmético-lógica,** que efectúa operaciones con los datos. La CPU también incluye un conjunto de registros y un conjunto de líneas, los buses, que permiten la comunicación con el exterior.

A este sistema también se le denomina **CPU** o Unidad Central de Proceso y constituye el núcleo operacional de un sistema más general: el ***computador*** *(o microcomputador).*

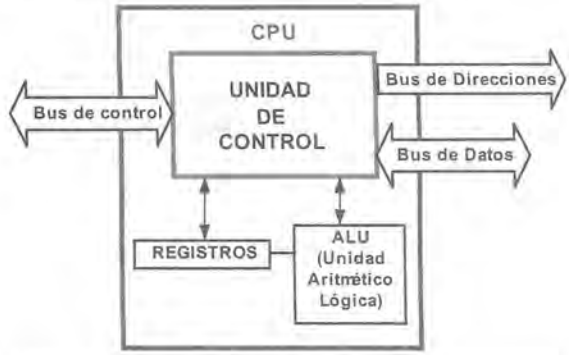


Figura 24.1. Diagrama de bloques simplificado de un procesador digital o CPU (Unidad Central de Proceso) compuesto por la unidad de control y la ALU.

Con anterioridad a la aparición de los microprocesadores la CPU de los sistemas programables se implementaba mediante varios circuitos integrados y componentes discretos. Con el microprocesador se dio un paso importante no sólo en la reducción de tamaño, que abrió considerablemente el abanico de aplicaciones, sino también en la potencia, coste y fiabilidad del sistema inherente al proceso de integración.

Un microprocesador no se utiliza de forma aislada y necesita de otros circuitos de apoyo para configurar un sistema computador. Entre los dispositivos externos están la **unidad de memoria** donde se almacena el programa y los datos y la **unidad de entrada/salida** que permite la transferencia de información con el exterior desde y hacia el computador. La memoria incluye la memoria ROM (*Read Only Memory*) o memoria de sólo lectura y la RAM (*Random Access Memory*) o memoria de lectura/escritura. El sistema de E/S adopta la forma de puertos, serie o paralelo que se comunican con variables digitales de proceso, con unidades de almacenamiento masivo o con otros sistemas periféricos: terminales, pantallas, tarjetas de captura de datos, etc.

La figura 24.2 muestra la configuración simplificada de un sistema computador típico basado en un microprocesador.

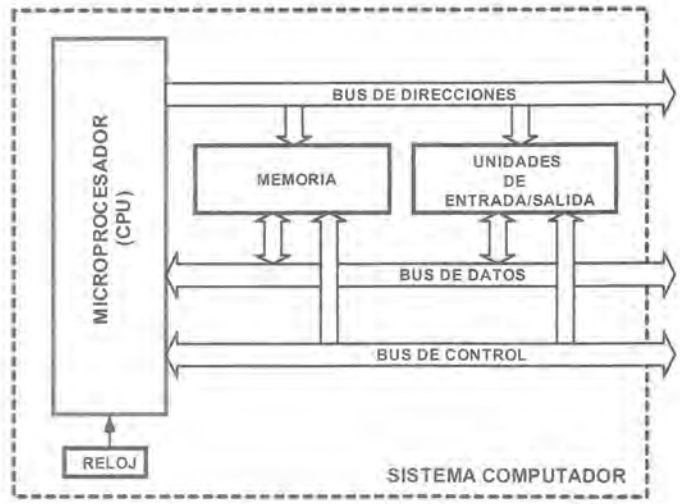


Figura 24.2. Diagrama de bloques de un sistema computador.

Como nexo de unión entre las distintas unidades del sistema se encuentra un conjunto de líneas denominadas buses (se corresponden con los buses señalados en la CPU) que se clasifican por el tipo de señales que contienen. El bus de direcciones es un bus unidireccional que permite seleccionar una determinada posición de la memoria o dispositivo de E/S, el bus de datos permite la transferencia de información entre los módulos y el de control contiene señales para la gestión y sincronización del sistema.

El aspecto más relevante de los microprocesadores es su carácter de sistema programable de **propósito general**, es decir, se concibe como un dispositivo universal capaz de satisfacer necesidades muy diversas, vinculadas generalmente con el **procesamiento de datos**. Este carácter de dispositivo de propósito general influye decisivamente en su estructura y repertorio de instrucciones. Por ejemplo, la mayoría de los μP disponen de un hardware muy sofisticado para el direccionamiento y la gestión de la memoria interna y externa.

Microcontroladores (μC)

Los sistemas programables basados en microprocesadores se hicieron muy populares en el ámbito de la gestión y también en el científico y el técnico, con configuraciones muy diversas que incluían sistemas com-

plejos de propósito general y sistemas más sencillos, de menores prestaciones, y dedicados a una tarea más específica que resolvían un problema concreto en la industria (generalmente relacionado con el control).

El desarrollo de los microprocesadores en el contexto industrial fue espectacular y pronto surgió la necesidad de un dispositivo similar a él pero que estuviese especialmente acondicionado para el trabajo en este entorno. En el contexto del control los requerimientos de procesamiento y acceso a datos no son tan acuciantes y los recursos de almacenamiento tampoco (no es necesaria tanta memoria, ni el tipo de periféricos que demandan las aplicaciones de procesamiento de datos). Más bien, la necesidad se orienta hacia una administración eficiente de los eventos o sucesos que surgen en un proceso para realizar su control. Un **microcontrolador** (μC) trata de responder a esta necesidad. Está basado, como el μP, en un circuito integrado, pero incluye en una misma pastilla de silicio gran parte de la funcionalidad que le permite actuar como un **sistema computador autónomo** adaptado a tareas de control y con una capacidad especial para conectarse directamente con el proceso. Así, un microcontrolador puede integrar en un mismo chip: memoria, circuitos temporizadores que permiten medir intervalos de tiempo o contar sucesos, convertidores A/D y D/A para la lectura y control de las variables del proceso, puertos de comunicación serie y paralelo, salidas y entradas digitales para el control de dispositivos discretos, etc.

La figura 24.3 muestra la estructura interna de un microcontrolador genérico donde se destaca su capacidad para capturar y actuar sobre las variables de proceso a través de los convertidores A/D y D/A y de los puertos de E/S digitales. La memoria se descompone generalmente en dos bloques: la memoria de programa, que almacena las instrucciones y la memoria de datos, que almacena las variables.

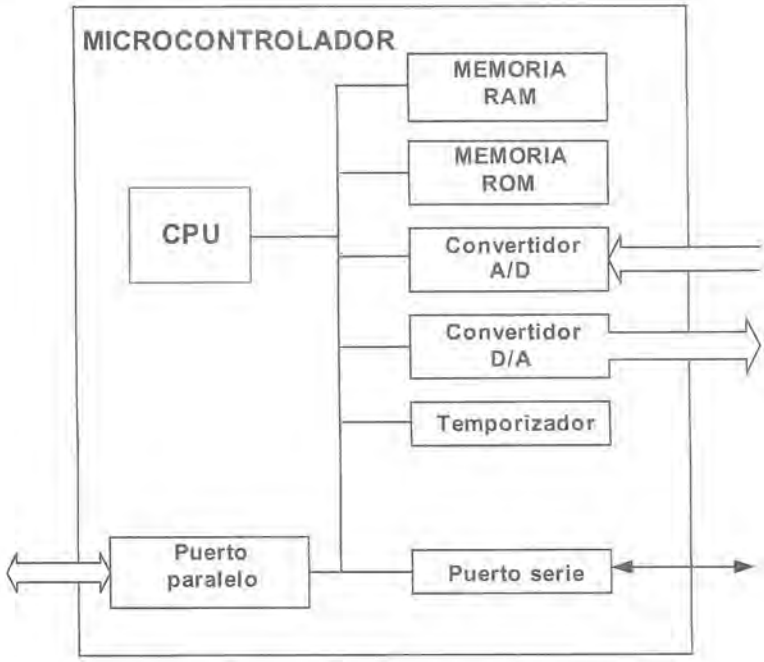


Figura 24.3. Diagrama de bloques de un microcontrolador.

Los μC pueden realizar desde tareas muy sencillas, obedeciendo a simples algoritmos de control hasta tareas muy complejas, sobre todo con las nuevas generaciones de microcontroladores que disponen de mucha potencia y funcionalidad. En general, suelen resolver un problema de control adaptándose al hardware que pretendan controlar, no permitiendo al usuario modificar su programa interno una vez que se ha programado para una función específica (sistemas empotrados).

La responsabilidad del éxito de los µC nace de la flexibilidad que aportan en el diseño de cualquier sistema electrónico y de su coste, aunque también hay que añadir su alta fiabilidad al reducir, gracias a la integración, el número de componentes discretos susceptibles de fallo. Además, son dispositivos pequeños y con bajo consumo que los convierte en idóneos para aplicaciones portátiles. Hoy día, su campo de aplicación es muy extenso y podemos encontrarlos en cualquier sistema de medida o circuito en que se requiera calcular o controlar. Los fabricantes suelen disponer de familias de microcontroladores con diversas versiones para adaptarse a cada necesidad. Algunos ejemplos de aplicación son el control de procesos, la industria del automóvil (climatización, control de inyección), los electrodomésticos (lavadoras, hornos, frigoríficos), la electrónica de consumo (cámaras de vídeo, televisores, vídeos), los equipos informáticos (ratones, impresoras), la instrumentación (equipos de medida, sensores inteligentes) y un largo etcétera.

Características de los microcontroladores

Los microcontroladores presentan, en general, una arquitectura interna de tipo **Harvard** con un espacio de memoria para datos y otro para el código, físicamente separados y con un sistema de buses dedicados a cada tipo de memoria (figura 24.4a). El objetivo es permitir el acceso simultáneo a las dos memorias para aumentar el rendimiento. La arquitectura de los computadores convencionales es de tipo **Von Newman**, es decir, los datos y el código comparten la misma área de memoria por lo que existe un sistema de bus único para el acceso tanto a datos como a código (figura 24.4b).

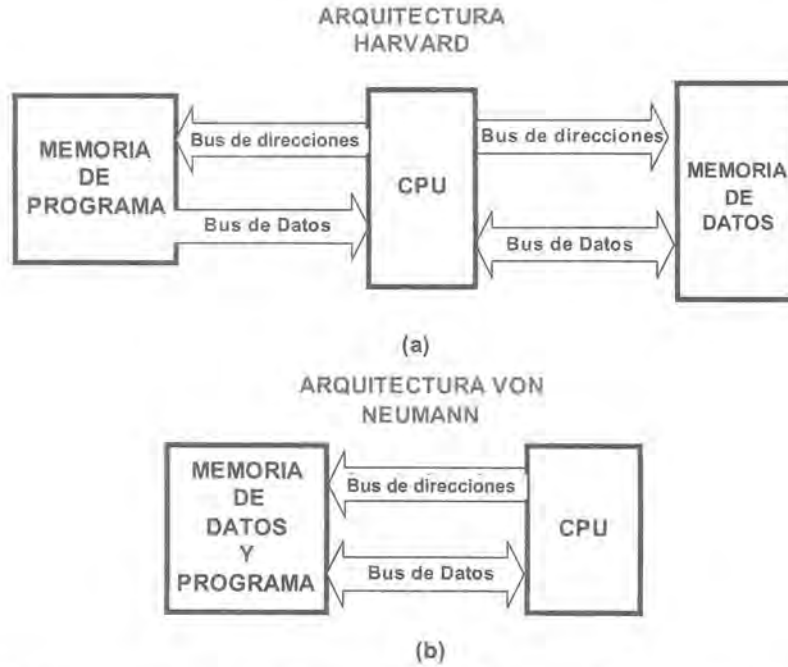


Figura 24.4. (a) Arquitectura Harvard con dos espacios de memoria diferentes para código y datos y con un doble bus; (b) arquitectura Von Neumann, con un espacio de memoria compartido para código y datos.

En cuanto a la estructura de las instrucciones, un gran número de microcontroladores son de tipo **CISC** (*Complex Instruction Set Computer*). Una instrucción de este tipo se convierte en una secuencia de microinstrucciones que deben completarse una a una para completar su ejecución consumiendo varios ciclos de reloj.

La tendencia actual es hacia los dispositivos de tipo **RISC** (*Reduced Instruction Set Computer*) que disponen de un conjunto de instrucciones muy reducido y que hacen tareas muy elementales pero que se ejecutan en un solo ciclo de reloj. El modelo RISC permite generar un código más optimizado aunque puedan requerirse varias instrucciones para realizar lo mismo que una instrucción de tipo CISC. Desde el punto de vista del compilador, la generación de código para un RISC es más eficiente. Además, como las instrucciones son sencillas y pocas, la unidad de control es más simple, ocupa menos silicio y permite la integración de mayor número de periféricos pudiendo trabajar a frecuencias superiores que los CISC y consumiendo menos energía.

Criterios de selección de microcontroladores

En la selección de un microcontrolador intervienen diversos factores que en última instancia dependen del tipo de aplicación a que vaya destinado. En general, los factores de decisión están relacionados con la capacidad de almacenamiento, capacidad funcional (número de temporizadores, contadores, número de puertos de E/S, número de canales A/D y D/A, etc.) y conectividad (capacidad de comunicación). No obstante, deben tenerse presentes otros factores como el consumo (muy importante en aplicaciones portátiles), el tamaño, el tipo de encapsulado, las herramientas de desarrollo, la disponibilidad de utilidades de software, de herramientas de depuración, el soporte técnico del fabricante, etc. A continuación, describiremos los factores de decisión más importantes en la selección de un microcontrolador:

Ancho de palabra: este término se refiere al número de bits de información que circula por su bus de datos (suele coincidir con el tamaño de sus registros internos). Los microcontroladores se clasifican según este criterio en μC de 4, 8, 16 y 32 bits. En general, el número de bits necesario para nuestra aplicación está relacionado con la precisión y la potencia que busquemos en nuestro sistema. En la actualidad el mercado *está dominado por los micros de 8 bits. Los μC de 16 y 32 bits suelen utilizarse en aplicaciones más exigentes en cuanto a velocidad y funcionalidad, por ejemplo, en aplicaciones de control donde el tiempo de respuesta* del sistema sea crítico (aplicaciones en tiempo real). Los μC de gama alta (32 bits generalmente) suelen incorporar un mayor número de prestaciones tanto computacionales (multiplicadores, etc.) como funcionales (convertidores D/A, generadores de señales, interfaces de comunicación avanzados, etc.).

Memoria: este parámetro es un elemento muy importante en la selección del dispositivo ya que debemos elegir uno con el espacio suficiente de memoria para almacenar tanto el código de programa (memoria de programa) como las variables (memoria de datos). La memoria interna que proporcionan los μC no suele ser muy extensa aunque en algunas versiones es posible su ampliación mediante el uso de memoria externa. En el caso de utilizar memoria externa debe considerarse la sobrecarga en el acceso, más lento que si se recurre exclusivamente a memoria interna. Las tecnologías de memoria que se utilizan en los μC son muy diversas. El tipo de memoria más popular para almacenar el programa es la memoria **EPROM** (*Erasable Programmable Read Only Memory*) que es una memoria no volátil programable mediante programadores especiales y que puede borrarse exponiéndola a una fuente de luz ultravioleta a través de una ventana que presenta el circuito. Proporciona poca versatilidad ya que el μC debe extraerse de su zócalo para grabarse y el proceso de grabado/borrado es lento. Otra alternativa es la memoria **EEPROM** (*Electrically Erasable Programmable Read Only Memory*) similar a la EPROM pero que es borrable eléctricamente, lo que simplifica el proceso de borrado. Sin embargo, es una memoria lenta y de coste elevado. La memoria **OTP** (*One Time Programmable*), es una memoria similar a la EPROM pero que sólo puede grabarse una vez (no dispone de ventana para borrado). Esta memoria es más económica que la EPROM y puede utilizarse una vez que nuestro programa ya esté depurado y tengamos la versión final, siempre que no sean muchas las unidades a producir. La memoria **Flash** es una memoria muy utilizada en la actualidad en sustitución de la EEPROM, sobre todo cuando es necesario disponer de un gran volumen de datos. Puede ser grabada y borrada (eléctricamente) en bloques, en vez de en bytes como la EEPROM y permite su grabación en el propio circuito de aplicación (*in circuit*). Finalmente, nuestro código puede residir en memoria de tipo **ROM** (*Read Only Me-*

mory) que es grabada por el fabricante (no por el usuario), siendo rentable cuando tenemos previsto un gran volumen de producción.

La memoria destinada a los datos es una memoria RAM volátil, generalmente estática (SRAM). El tamaño de la memoria necesaria dependerá del tipo de aplicación, debiendo tener en cuenta el número de funciones de nuestro programa y su extensión.

Convertidores A/D: si es necesario capturar una señal analógica externa debemos considerar un μC con esta prestación incorporada. El tipo de convertidor más usual es el de aproximaciones sucesivas. En general, suelen disponer de varios canales, por lo común multiplexados (4, 8, 16), con diversas resoluciones (8, 12, 16 bits) y no muy rápidos. Es interesante considerar los algoritmos disponibles de exploración de los canales y los niveles de tensión que se admiten. En algunos casos, resulta interesante que el μC permita fijar el alcance de tensión de entrada del conversor a través de referencias de tensión externas, de forma que no se limite a los niveles de alimentación del μC (0-5 V generalmente).

Convertidores D/A: si es necesario proporcionar una señal analógica de salida podemos seleccionar un μC que incluya esta característica. Las tensiones y corrientes de salida son limitadas y también disponen de varias resoluciones.

Temporizadores/contador: este recurso permite la medición del tiempo que transcurre entre dos sucesos o contar el número de veces que se produce un determinado evento. La resolución de la temporización depende de la frecuencia de reloj utilizado por el sistema y la capacidad de temporización, del tamaño del registro destinado a este propósito; el temporizador (*timer)* suele tener un tamaño de 8, 16 ó 32 bits. En la versión de contador dispone de una entrada activa por flanco que incrementa un registro interno en cada flanco activo. En general, el hardware asociado al temporizador es muy versátil: se permiten escalados, recargas automáticas con un valor determinado cuando se produce un evento especial, activación de interrupciones cuando la temporización llega a su final, etc. Es importante considerar la posibilidad de disponer de dos o más temporizadores independientes.

Temporizador *watchdog*: este recurso permite la reinicialización del programa de forma automática en caso de que este se bloquee por cualquier circunstancia. Es muy útil cuando el μC trabaja en entornos ruidosos o en los casos de sistemas que no están sometidos a vigilancia. En algunos μC existe también la posibilidad de detectar caídas de tensión en la alimentación provocando un reset automático para evitar inestabilidades (*Brown-out Reset*).

Sistema de interrupciones: las interrupciones constituyen un método asíncrono de interrumpir el flujo de un programa. Casi todos los μC incorporan esta posibilidad mediante entradas externas o vinculadas a sucesos particulares: nivel en una entrada, flanco en una entrada, *timer* a cero, carácter recibido en el puerto serie, etc. También es importante considerar si existe algún mecanismo de prioridades en la interrupciones.

Comunicación: cuando pretendemos comunicar el μC con otros periféricos (computadores, displays, impresoras, etc.) es interesante que se disponga de puertos de comunicación. Lo más habitual es que se incluyan dispositivos de comunicación serie asíncrona y síncrona, con hardware especialmente dedicado a esta tarea (*UART, Universal Asyncronous Receiver Transmiter*). Aunque puede implementarse un puerto serie en software, el hecho de que exista el recurso físicamente permite disponer de más memoria para nuestro programa y, además, la comunicación es más eficiente.

Algunos μC disponen de conectividad con buses de comunicación avanzados (CAN, I^2C, etc.) debido a la fuerte implantación de los microcontroladores en el control distribuido. En algunos casos, puede incrementarse la memoria del microcontrolador por medio de memoria externa que se conecta a través de estos buses por lo que es interesante su consideración.

Puertos de entrada/salida: son puertos disponibles para la lectura y escritura de información digital. Un puerto está compuesto por 8 bits individuales que pueden configurarse como entrada o salida siendo

importante considerar el número de puertos disponibles. Algunos puertos pueden controlarse de forma individual, es decir, puede indicarse bit a bit cuál es de entrada y cuál de salida. También es posible la implementación de un protocolo de comunicación sobre un puerto para comunicarse con un periférico externo o con otro microprocesador. El número de puertos necesario vendrá determinado por la cantidad de señales digitales que deseemos controlar en nuestra aplicación. Debe considerarse que algunos puertos pueden estar dedicados a tareas específicas o multiplexan dos o más funciones por el mismo pin. También es interesante la posibilidad de que disponga de pines con salidas en colector abierto para poder fijar externamente el nivel de tensión. Los niveles de corriente que suelen manejar no son elevados aunque puede seleccionarse μC que disponen de pines especiales que proporcionan mayores niveles de corriente. Los niveles de tensión de los pines suelen ser TTL.

Unidades de generación de señales especiales: dependiendo de la aplicación, algunos modelos disponen de hardware especializado para la generación de señales como trenes de pulsos, modulación por anchura de pulsos (PWM), etc.

Herramientas de desarrollo: es muy importante considerar las herramientas disponibles ya que pueden reducir considerablemente el ciclo de desarrollo de nuestra aplicación. En general, las herramientas de software que se utilizan son: ensambladores (convierte lenguaje ensamblador a código máquina), enlazadores (junta los módulos que componen una aplicación en un único módulo resolviendo las referencias entre módulos, variables y llamadas de subrutinas, etc.), compiladores (permiten utilizar lenguajes de alto/medio nivel, como el C) y simuladores (permiten verificar el funcionamiento del programa desarrollado sin necesidad de introducirlo físicamente en el μC). Esta herramienta es muy útil en la depuración ya que permite la ejecución paso a paso, el establecimiento de puntos de ruptura, la visualización de los registros internos, etc.

También es importante que el μC disponga de herramientas hardware de depuración como los **emuladores** (hardware que ejecuta físicamente nuestro programa pero dentro de un entorno controlado que permite *verificar y depurar el sistema desarrollado). Otra opción son las* ***tarjetas de evaluación*** *destinadas a la evaluación* de las prestaciones del μC y que pueden servir para el desarrollo de prototipos. Las tarjetas de evaluación suelen incorporar un hardware auxiliar (pulsadores, convertidores de niveles, puertos serie, zócalos de ampliación memoria, displays, etc.) para dotarlas de la mayor versatilidad posible.

Consumo de energía: es importante considerar el consumo del μC, sobre todo en aplicaciones portátiles. Algunos modelos permiten el funcionamiento en un estado de bajo consumo cuando la actividad de procesamiento cesa por algún motivo. También es importante considerar la posibilidad de trabajar con tensiones inferiores a los niveles de tensión tradicionales (5 V TTL).

Procesadores Digitales de Señal (DSP)

Un DSP (***Digital Signal Processors***, Procesador Digital de Señal) es un procesador digital cuyo hardware y conjunto de instrucciones está optimizado para la implementación eficiente de aplicaciones de **procesamiento numérico intensivo a alta velocidad**. Este tipo de procesadores ejecutan, generalmente, los algoritmos típicos de procesamiento digital de señal (***Digital Signal Processing***) como el filtrado digital y el análisis espectral[1], entre otros. Este tipo de algoritmos se requieren a menudo en sistemas que procesan señales analógicas en **tiempo real**. Con tiempo real queremos indicar que el sistema debe dar una respuesta en un tiempo establecido y predecible. Por ejemplo, un escenario habitual es un sistema que adquiriera datos muestreando una señal analógica para someterla a un determinado procesamiento, por ejemplo, el filtrado. Esta aplicación exige que se realicen todos los cálculos antes de que llegue la siguiente muestra. Si el procesamiento de cálculo es muy complejo, deberán utilizarse procesadores especializados para esta tarea porque, de otra forma, se perderían muestras.

[1] En la bibliografía se proporcionan referencias a textos especializados en este tipo de algoritmos.

Las aplicaciones de los DSP son muy variadas: procesamiento de señales de audio (compresión de audio, filtrado, TV digital), procesado de habla (reconocimiento de voz, compresión de voz, grabadores digitales de habla), comunicaciones (teléfonos móviles, módems, fax, GPS), multimedia (compresión de vídeo, audio, imagen digital), control industrial (cancelación de ruido, control de motores, fuentes de alimentación) y un largo etcétera.

También tienen diversas aplicaciones en el campo de la instrumentación. Las funciones básicas que se requieren de un DSP en este campo son: FFT (transformada rápida de Fourier), filtrado, síntesis de ondas, filtros adaptativos, cálculos numéricos a alta velocidad, etc. Las aplicaciones incluyen: equipos de test y medida, análisis de vibraciones, tarjetas de E/S para PCs (tarjetas de adquisición de datos), analizadores de señal, generadores de funciones, etc.

Para darnos una idea del tipo de algoritmos y operaciones que se realizan en un DSP vamos a considerar una aplicación muy frecuente como es el **filtrado digital**. En la figura 24.5 se muestra el diagrama bloques de un sistema basado en un DSP.

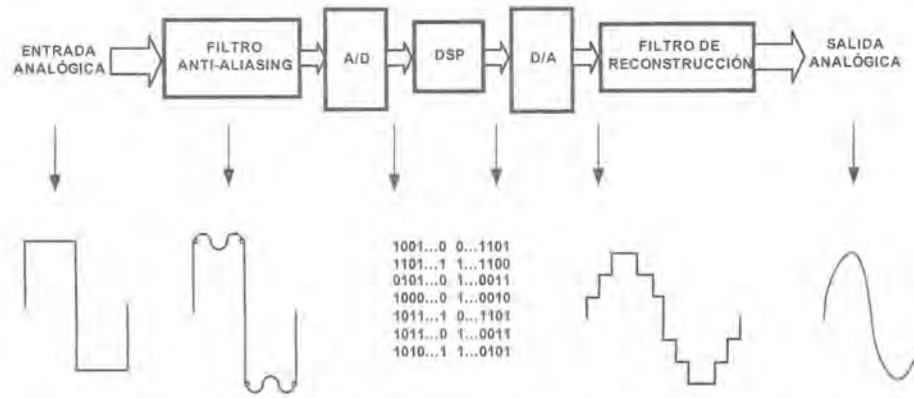


Figura 24.5. Sistema de adquisición basado en un DSP.

La señal analógica de entrada es una señal cuadrada que deseamos filtrar para obtener a la salida una señal senoidal como la mostrada en la figura 24.5. En general, existen dos métodos para adquirir datos: **procesamiento continuo** o por muestras y el **procesamiento por lotes** (*frames*). En sistemas basados en muestreo, como el propuesto, en cada pulso de reloj se obtiene un dato, se procesa y se obtiene la salida correspondiente de forma continua. En los sistemas basados en *frames*, como los analizadores de espectros, se adquieren los datos por bloques enteros que después se procesan en conjunto y se da la salida. En nuestro caso, las señales que provienen del convertidor A/D llegan al DSP como un tren de muestras individuales a la frecuencia de muestreo del convertidor A/D que es función del ancho de banda de la señal de entrada. Para evitar el submuestreo se somete a la señal a un filtrado previo analógico (filtro *antialising*). El DSP realiza el algoritmo de filtrado a través de un proceso numérico y genera una salida que se envía hacia el convertidor D/A para obtener una señal analógica que se reconstruye con la ayuda de un filtro pasa-bajo. Opcionalmente, el DSP puede comunicarse con otros equipos para enviar o recibir datos o información de control provenientes de un computador.

Por regla general, para realizar el filtrado en tiempo real, el DSP debe realizar todos los cálculos y operaciones requeridas para procesar cada muestra antes de que llegue la siguiente. El tiempo entre muestras es el que dispone el DSP para realizar todos los cálculos.

EJERCICIO 24.1

Una señal de audio se desea filtrar realizando un procesamiento continuo. Se muestrea a 48 kHz. Calcule el tiempo que se dispone para realizar el procesamiento. ¿Cuánto tiempo se dispondría si se adquiere en frames de 1024 muestras cada uno?

SOLUCIÓN:

El tiempo entre muestras es el tiempo que disponemos para realizar el procesamiento:

$T_d = 1/48000 = 20{,}833$ µs. Si se adquiere en frames, se dispone de: $1024 \cdot 20{,}833$ µs = 21,33 ms

Un ejemplo del tipo de procesamiento que se realiza en el filtrado se muestra en la figura 24.6, donde se refleja la función de transferencia de un filtro digital muy utilizado: el filtro FIR (*Finite Impulsive Response*).

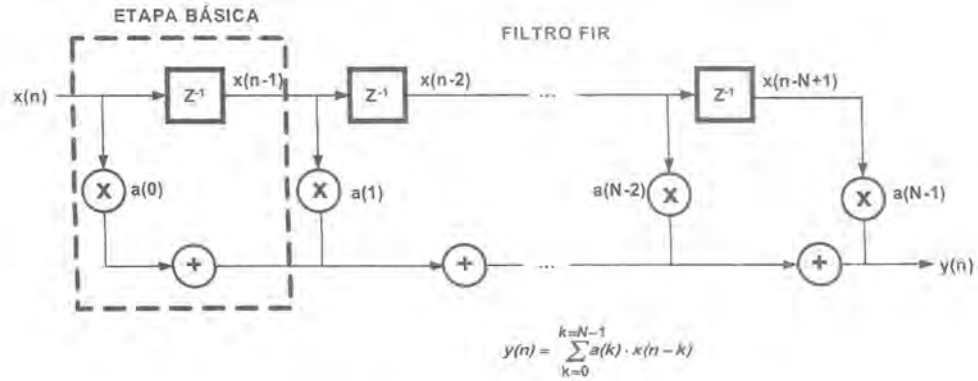


Figura 24.6. Estructura de un filtro FIR.

El término x(i) representa las muestras de entrada, y(i) representa las muestras de salida, a(i) son los coeficientes del filtro, N es el orden del filtro y Z^{-1} es un operador que aplicado sobre una muestra obtiene la muestra anterior.

Las señales x(n) representan un conjunto de valores discretos obtenidos por muestreo de la señal analógica de entrada. Los bloques cuadrados de la figura representan retardos unitarios, es decir, su salida es el valor de la entrada pero retardada en un periodo de muestreo, es decir, obtiene la muestra anterior a la aplicada a su entrada. El procesamiento que se debe realizar para obtener la salida consiste en considerar la muestra actual x(n) y un conjunto de muestras anteriores, x(n-1), x(n-2),... e ir multiplicándolas por los coeficientes del filtro, a(0), a(1), a(2), ... Después se suman para calcular el resultado, y(n). Cada vez que llega una nueva muestra las demás se desplazan una posición hacia la derecha de manera que se vuelve a calcular la salida y así, sucesivamente. La aproximación a un filtro ideal consiste en aplicar la función de transferencia con los coeficientes apropiados y el orden suficiente.

Como se observa en este ejemplo, y es algo que se mantiene en numerosos algoritmos de procesamiento digital de señal, se realizan operaciones matemáticas de gran complejidad y sofisticación a través de operaciones sencillas, básicamente multiplicaciones y sumas acumuladas. Por ejemplo, la función básica del algoritmo de filtrado consiste en una repetición sistemática de una multiplicación y suma acumulada del tipo $s = r + a \cdot x$ que se repite un número N de veces, según el orden del filtro.

El DSP está especialmente acondicionado para realizar este tipo de operaciones y repetirlas sistemáticamente de forma eficiente, generalmente utilizando un solo ciclo de reloj. Comparado con un microprocesador genérico o con un microcontrolador, el DSP presenta unas diferencias relacionadas básicamente con su capacidad numérica de computación. Los µP tienen un carácter más universal, no están pensados para ninguna aplicación específica y menos para el procesamiento numérico intensivo; los µC están más dirigidos hacia el control. Ambos requieren varios ciclos de reloj para completar el tipo de operaciones que se mencionan lo que puede resultar muy costoso en tiempo, sobre todo cuando el proceso se hace muy repetitivo, algo habitual en los algoritmos de procesamiento digital de señal. El DSP posee unidades funcionales muy potentes (ALU, multilplicador, registros, desplazadores de bits, procesadores de E/S, puertos de alta velocidad, etc.) explotándose al máximo el paralelismo entre ellas. También dispone de recursos hardware para la gestión de estructuras de datos y de control cuyo objeto es reducir el tiempo de ejecución de algoritmos críticos (*buffers* circulares, bucles optimizados, etc.). En los µP y µC se recurre al software para implementar estas tareas o, simplemente, no disponen de estos recursos.

A las capacidades de cálculo del DSP debe añadirse además su capacidad para realizar transferencias eficientes de datos procedentes del exterior, en general, disponiendo de puertos de alta velocidad y procesadores especializados en transferencias de E/S de tipo DMA (*Direct Memory Access*). Por ejemplo, en el algoritmo de filtrado no sólo debe de calcularse a gran velocidad, sino que también deben de proporcionarse los datos a un ritmo sostenido. A continuación, analizaremos estas características con más detalle.

Características de los DSPs

Un DSP típico incluye en la misma pastilla de silicio la unidad central de proceso, la memoria y, generalmente, un procesador especializado en operaciones de E/S. La figura 24.7 describe la arquitectura de un DSP genérico.

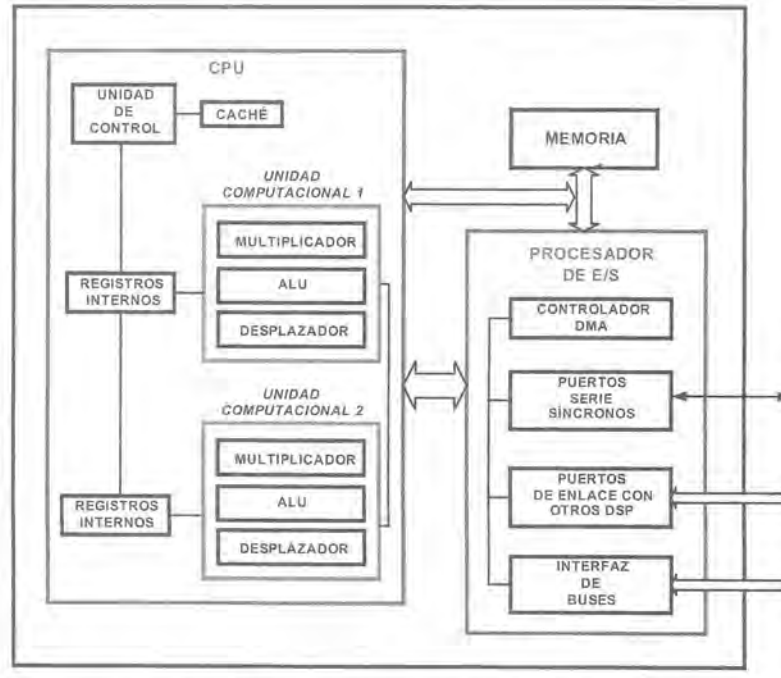


Figura 24.7. Arquitectura de un DSP. La CPU puede estar compuesta por una o varias unidades de cálculo que pueden trabajar en paralelo. El procesador de E/S permite la conexión con dispositivos externos y facilita la tarea de movimiento de grandes volúmenes de datos.

CPU

La CPU está integrada por la unidad de control y una o varias unidades funcionales (en este ejemplo, se consideran sólo dos) que incluyen: un multiplicador con capacidad para realizar multiplicaciones y sumas acumuladas (se suele denominar *MAC, Multipler/ACumulator*), un desplazador (*Barrel shifter*, que permite hacer desplazamientos de múltiples bits con una sola instrucción) y una ALU para realizar operaciones aritméticas (*suma/resta*) y *lógicas. Sobre todo, destaca el multiplicador (MAC) implementado totalmente en* hardware y con capacidad de realizar una **multiplicación y una suma acumulada en un ciclo de reloj**. Esta implementación no es habitual en μC (se recurre a software) pero es muy frecuente en los algoritmos de procesamiento digital como vimos en el caso del filtro. Las unidades de cálculo son muy rápidas y flexibles y además pueden paralelizarse, es decir, pueden realizar varias operaciones concurrentemente. Por ejemplo, se puede realizar una multiplicación y una suma acumulada (utilizando el multiplicador) y al mismo tiempo realizar una operación aritmética de suma y/o resta utilizando la ALU en la misma unidad funcional y todo en un solo ciclo de reloj (en procesadores de propósito general o microcontroladores se requerirían varios ciclos de reloj). Si existen varias unidades funcionales (como se muestra en la figura 24.7) se incrementa aún más el grado de paralelismo.

Los DSPs están muy optimizados para el cálculo numérico siendo su distinción esencial frente a los sistemas vistos anteriormente. Así, disponen de un repertorio de instrucciones aritméticas muy potente (1/x, 1/sqrt(x), mín, máx, rotación, desplazamiento, multiplicación, suma, multiplicación y suma combinada, etc.) y un extenso número de registros internos, incluyendo registros de tamaño extendido para permitir un margen dinámico adecuado en ciertas operaciones[2]. Una característica importante en los DSPs es el tipo de formato aritmético utilizado por el procesador; de hecho, constituye el criterio de clasificación más habitual. En general, se utilizan formatos de coma fija o flotante lo que da lugar a dos tipos de DSP. El formato de coma flotante proporciona un mayor margen dinámico (cociente entre el valor mayor y el menor valor representable) que el de coma fija. En la figura 24.8 se muestran los formatos de coma fija y coma flotante.

Figura 24.8. Formatos de coma fija y coma flotante. S corresponde al bit de signo, e al exponente y f representa los bits significativos del formato en coma fija.

EJERCICIO 24.2

Calcular el margen dinámico de un DSP de coma fija de 32 bits y uno de coma flotante de 32 bits siguiendo el formato de datos mostrado en la figura 24.8.

---

[2] Es muy habitual que los resultados de operaciones con operandos de 16 bits, por ejemplo, se acumulen en registros de 40 bits o más para evitar desbordamientos (*overflows*).

SOLUCIÓN:

El número máximo representable en coma fija es: $1-2^{-31}$ y el mínimo es: 2^{-31}. El cociente representa el margen dinámico: $1-2^{-31}/\ 2^{-31} = 2{,}147\ 10^9$ que se expresa en dB como: $20\log 2{,}147\ 10^9 = 187$ dB.

En el formato de precisión simple en coma flotante (estándar IEEE754/854) los números normales se representan como: $x = (-1)^s 2^{(e-127)}(1,m)$ donde m es la mantisa de 23 bits, e es el exponente con $1 \leq e \leq 254$ y s el bit de signo. El mínimo número representable es $\approx 2^{1-127}$ y el máximo es $\approx 2^{254-127}$ cuyo cociente resulta $1{,}447 \cdot 10^{76}$ y expresado en dB, $20\log 1{,}447 \cdot 10^{76} = 1523$ dB. Como se puede observar el margen dinámico es mucho más amplio en los DSPs de coma flotante.

Formando parte de la unidad de proceso de un DSP también suele incorporarse una memoria especial, denominada memoria caché, destinada al almacenamiento de las instrucciones más usadas con el objeto de aumentar las prestaciones del sistema. El acceso a esta memoria local es muy rápido y no se utilizan los buses que comunican con la memoria de datos y programa (figura 24.7) para transferir la instrucción a la unidad de control. De esta forma, los buses, tanto el de programa como el de datos, quedan libres para realizar un acceso a operandos al mismo tiempo que se inicia un ciclo de búsqueda de instrucción en la caché. Evidentemente, este proceso se interrumpe si la instrucción no se encontrase disponible en la memoria caché pues su tamaño es limitado. No obstante, muchos algoritmos de procesamiento digital de señal suelen estar constituidos por un conjunto de instrucciones no muy extenso que se repite de forma reiterada.

El rendimiento es esencial en un DSP y, por ello, se utilizan arquitecturas avanzadas de procesamiento. Una arquitectura muy habitual consiste en incorporar varias unidades de cálculo como se mostró en la figura 24.7. En la figura 24.9 se detalla el flujo que siguen los datos desde la memoria hasta las unidades funcionales. Como se aprecia, una misma instrucción se aplica sobre dos flujos de datos independientes. A esta arquitectura se le denomina ***SIMD*** *(Single Instruction Multiple Data)* o instrucción simple de datos múltiples.

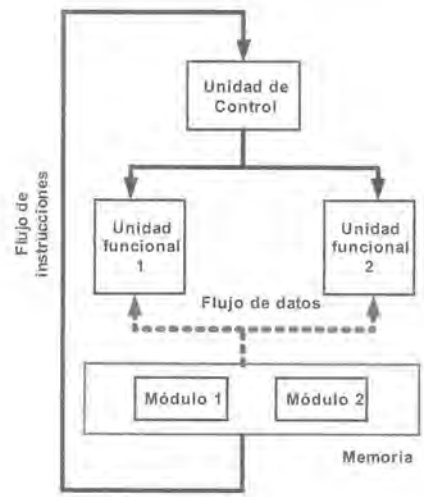


Figura 24.9. Estructura SIMD. Los datos llegan a las dos unidades funcionales que realizan las operaciones de forma simultánea.

Básicamente, los datos se descomponen en bloques que son procesados de forma paralela en cada unidad funcional, pero aplicando la misma instrucción. Esta técnica es muy habitual en operaciones sobre vectores

(se utilizan en operaciones como convoluciones o transformaciones). Otra alternativa para incrementar la capacidad de procesado es ejecutar más de una instrucción por ciclo de reloj. Esta alternativa viene de la mano de las arquitecturas **VLIW** (*Very Long Instruction Word*) basadas en una instrucción con un formato de palabra grande compuesta por varias instrucciones independientes que involucran a varias unidades funcionales cada una de ellas ejecutando una instrucción diferente.

Otras características importantes en los DSPs son los recursos que se proporcionan para implementar y administrar las estructuras de datos que habitualmente se utilizan en los algoritmos de procesamiento digital de señal como por ejemplo los *buffers circulares*[3]. Como el rendimiento es un elemento clave en estos sistemas, también suele utilizarse hardware especializado para el direccionamiento de datos. Por ejemplo, es muy habitual que existan generadores de direcciones a datos y a código independientes del secuenciador de la unidad de control. También se incrementa el rendimiento con la optimización en la ejecución de las estructuras repetitivas como bucles y saltos donde, en general, se evitan las pérdidas de tiempo asociadas a la comprobación de fines de bucle, cumplimiento de condiciones en bucles, redirección al inicio de bucles[4], etc. Esta característica es muy importante porque, como comprobamos en el ejemplo del filtrado, en muchos algoritmos se repiten sucesivamente las mismas operaciones de forma continua y un pequeño retardo en cada iteración puede suponer mucho tiempo de cálculo perdido.

Memoria

Los DSPs presentan generalmente una arquitectura Harvard con un bus dedicado a la memoria de programa y otro para la memoria de datos. Por regla general, la memoria de trabajo es de doble puerto y con varios

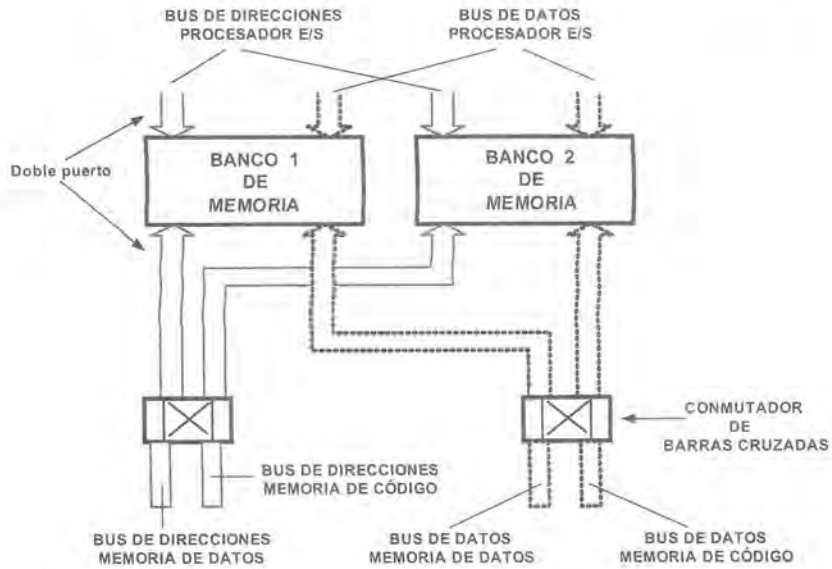


Figura 24.10. Ejemplo de la estructura de memoria interna de un DSP (ADSP21061 Analog Devices). El sistema de memoria se basa en un doble banco de memoria de doble puerto. Un puerto se dedica al procesador de E/S y el otro se comparte entre el bus de programa y el bus de datos mediante un conmutador de barras cruzadas.

[3] Estructuras basadas en segmentos circulares de memoria, cuyo mecanismo de direccionamiento se optimiza para aumentar la velocidad de ejecución.

[4] Esta característica se denomina *zero-looping overhead* o *sobrecarga nula en los bucles*.

bancos, permitiendo accesos simultáneos tanto por parte de la CPU (por ejemplo, acceso a dos operandos simultáneamente) como por el procesador de E/S (figura 24.10). En los μP el hecho de que mantengan una estructura de bus único (von Neumann) penaliza los accesos a memoria. En general, la memoria interna que proporciona un DSP suele ser de mayor dimensión que la que se integran habitualmente en los μC y puede llegar a configurarse por el usuario según sus necesidades, asignando qué cantidad destina a código y a datos (en los μC esta división es fija).

Procesador de E/S

La conectividad con el exterior es otro factor clave cuya responsabilidad reside en el módulo etiquetado como procesador de E/S (figura 24.7). El mantenimiento de un flujo de datos sostenidos es una exigencia habitual en muchas aplicaciones. Por ejemplo, datos que pueden venir del convertidor A/D deben transferirse a la memoria para ser procesados. Los DSPs suelen incorporar para esta tarea canales de comunicación con el exterior de alta velocidad garantizando velocidades de transferencia sostenidas y no sólo a picos. Generalmente, integran puertos serie síncronos de alta velocidad para la comunicación con dispositivos externos. Estos puertos están conectados con los canales del DMA. El DMA es un procesador de E/S que permite la transferencia de bloques de datos entre la memoria y los dispositivos externos sin la intervención directa de la CPU que puede seguir realizando cálculos o, incluso, acceder a memoria de forma simultánea a la transferencia de E/S (24.11). Este tipo de prestaciones no se integran en los μP ni en los μC aunque están disponibles en sistemas basados en μP a través de circuitos independientes.

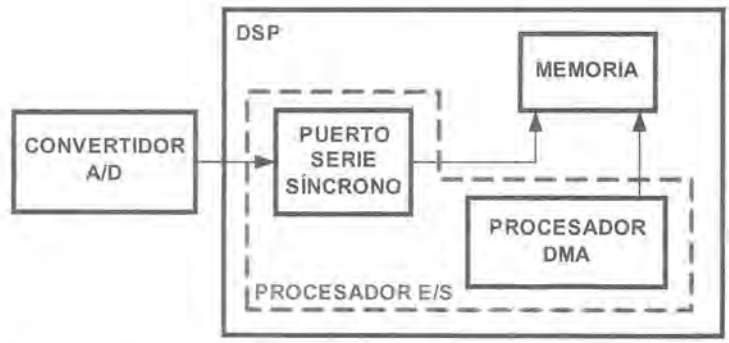


Figura 24.11. Transferencia de datos a través de un puerto síncrono mediante DMA.

En relación con la conectividad, también es importante la capacidad de interconexión de un DSP con otros DSPs para configurar sistemas multiprocesadores que trabajen en paralelo (figura 24.12). Esta configuración suele utilizarse en sistemas que demanden gran potencia de cálculo (radares, análisis sísmico, etc.). El método de funcionamiento consiste en repartir el trabajo entre varios procesadores siempre que el algoritmo permita paralelizarse. Existen varios métodos para realizar esta tarea. El más tradicional es enlazar los procesadores de manera que el flujo de datos pase de un DSP a otro, de forma similar al funcionamiento que se realiza en una cadena de montaje (estructura tipo *pipeline*). Otro método consiste en conectar los procesadores formando una matriz, una estrella u otra configuración geométrica donde la comunicación se establece a través de puertos especializados. La tercera posibilidad consiste en utilizar un espacio de memoria compartida al que todos los procesadores tienen acceso. Esta posibilidad presenta la limitación de que en general sólo un procesador puede utilizar la memoria compartida al mismo tiempo.

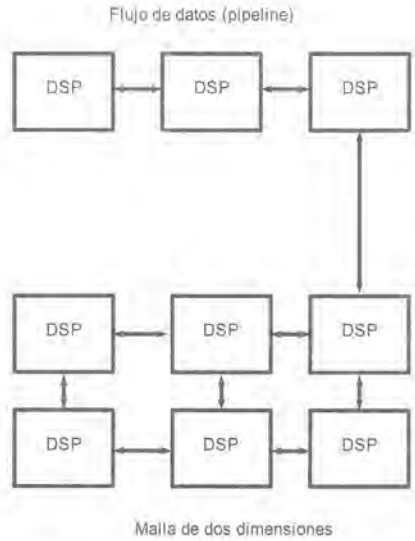


Figura 24.12. Sistemas multiprocesadores, conexión en *pipeline* o en malla de dos dimensiones.

Finalmente, debe considerarse la posibilidad de conexión del DSP con otros dispositivos, como computadores, microprocesadores, microcontroladores o memoria externa a través del bloque etiquetado como interfaz de bus en la figura 24.7.

En la actualidad se ha llegado al convencimiento de que el aumento de la velocidad en la computación dependerá de la explotación del paralelismo entre unidades funcionales ya que existen límites físicos para incrementar la velocidad de reloj de forma continuada. El problema que se plantea es cómo aprovechar el paralelismo de la forma más eficiente posible y aquí el desarrollo de los compiladores juega un papel clave. En el mundo del DSP el diseño de algoritmos eficientes que aprovechen el hardware al máximo es un asunto de importancia capital. En realidad, el software es el componente más crítico en el diseño de un sistema de este tipo.

Hoy en día microcontroladores y DSPs tienden a una cierta convergencia porque cada vez más μC disponen de unidades de cálculo más potentes. Así, surgen los sistemas híbridos que combinan las prestaciones de ambos. El grado actual de integración también está llevando hacia el diseño de circuitos en los que se integran varios DSPs en el mismo chip.

Criterios de selección de un DSP

En lo que se refiere al procesamiento digital de la señal la elección del dispositivo adecuado es una de las decisiones más críticas. A continuación, analizaremos ciertos criterios que deben considerarse en la elección de un DSP.

Velocidad de ejecución: este criterio es básico a la hora de seleccionar un DSP. Para determinar los requerimientos de velocidad del DSP deben establecerse los requerimientos de ancho de banda de la señal de entrada y fijar la frecuencia de muestreo. También debe determinarse si se va realizar el procesamiento entre muestras como el ejemplo de filtrado, o de un paquete completo de datos (típico en FFT) ya que se dispone entonces de más tiempo. La velocidad de ejecución necesaria en el DSP dependerá de la complejidad del algoritmo y del porcentaje de tiempo que se dedique a procesarlo. Esta velocidad puede medirse de diversas formas:

- **MFLOPS**: *Million Floating-Point Operations Per Second.* (millones de operaciones en coma flotante por segundo). Se refiere al número de operaciones en coma flotante (multiplicaciones, sumas, restas, etc.) que puede realizar en un segundo. Es importante contrastar si lo que se especifica se trata de picos o de valores sostenidos.
- **MOPS**: *Million Operations Per Second.* (millones de operaciones que puede ejecutar el procesador en un segundo). Incluye operaciones de entrada/salida y de transferencia de datos (DMA), además de las matemáticas. Ofrece una visión de las capacidades de cálculo y de E/S.
- **MIPS**: *Million Instruction Per Second.* (millones de instrucciones por segundo). Esta medida es muy habitual y se calcula como el inverso del tiempo del ciclo de instrucción dividido por un millón. Casi todos los DSPs ejecutan una instrucción por ciclo de máquina aunque debe considerarse que el rendimiento final va depender de qué es lo que puede ejecutar cada instrucción. En general, el hardware especial de ciertos procesadores está concebido con un alto grado de paralelismo de manera que se pueden realizar en una misma instrucción varias operaciones simultáneamente, según el número de unidades funcionales de que disponga. Así, un procesador con 10 MIPS y capaz de realizar 8 operaciones en paralelo por cada instrucción presenta básicamente el mismo rendimiento que un procesador de 40 MIPS y que sólo realice 2 operaciones por instrucción. Por tanto, hay que tener precaución con este parámetro.
- **MBPS**: *Mega Bytes Per Second.* (megabytes por segundo). Este parámetro suele usarse para evaluar las prestaciones de ancho de banda de un bus o un puerto de entrada/salida.

Estos valores pueden servirnos para ubicar cada DSP dentro de una determinada categoría, pero a veces no dicen mucho de lo que realmente es capaz de hacer.

Un método de estimación de velocidad más práctico y realista puede ser definir un algoritmo típico como por ejemplo un filtro FIR (de un determinado orden) e implementarlo en distintos DSPs para comparar prestaciones.

EJERCICIO 24.3

Una señal de audio se muestrea a 48 kHz. El algoritmo de procesamiento requiere ejecutar 50 operaciones para calcular el resultado. ¿Qué velocidad se requiere en el DSP?

SOLUCIÓN:

A la frecuencia de muestreo le corresponde un periodo entre muestras de 20,833 µs. Si pretendemos procesar los datos en tiempo real, la velocidad del DSP que se requiere es, expresado en millones de operaciones por segundo (MOPS):

$$\text{velocidad DSP} = \text{operaciones/intervalo de muestreo} = 50/20{,}833\ \mu s = 2{,}4\ \text{MOPS}$$

Así, si el DSP dispone de todo el tiempo entre muestras para realizar el algoritmo (aunque lo habitual es que además tenga que realizar más cosas), un procesador con un rendimiento superior a 2,4 MOPS puede ser buen candidato. Generalmente, la velocidad en los procesadores se expresa en MOPS o en MIPS (millones de instrucciones por segundo). Recuerde que estos dos parámetros no indican lo mismo.

Arquitectura de memoria: la disponibilidad de memoria interna es otro factor crítico. Es deseable disponer de un sistema flexible para poder alojar código y datos según determine el usuario. Lo ideal sería disponer de la suficiente memoria interna para dar cabida a todo el programa, código y datos ya que se au-

menta la velocidad de proceso; en caso contrario se hace necesario realizar accesos externos. Por ejemplo, el ADSP21160[5] dispone de dos bloques de memoria independientes de 512 Kbytes que permiten realizar accesos simultáneos a datos. Este DSP presenta una arquitectura del tipo SIMD (*Single Instruction Multiple Data*) donde una sola instrucción desencadena operaciones sobre distintos flujos de datos; para ello dispone de dos unidades de computación independientes que se conectan a cada bloque de memoria, lo que aumenta considerablemente las prestaciones. La presencia de una caché para las instrucciones es otro factor relevante: permite, por ejemplo, realizar el acceso a dos datos simultáneos que residen en los bloques de memoria y, de forma simultánea, el acceso a la instrucción que reside en la caché, todo ello en un solo ciclo de reloj. La disposición de un bloque extenso de caché, por ejemplo los 128 Kbytes en el TMS320C6203[6], permite la carga completa de bloques de código. En general, se habilita la caché durante los primeros accesos a memoria externa, de manera que una vez que residan en la caché pueden obtenerse prestaciones muy elevadas. Esta característica es vital en determinadas aplicaciones.

El modelo TMS320C6701[7] dispone de una arquitectura del tipo VLIW (*Very Long Instruccion Width*) de 256 bits con capacidad para programar 8 elementos de cálculo que pueden trabajar en paralelo. La memoria, en este caso, se organiza en palabras de 256 bits, ejecutándose una instrucción en cada ciclo. El rendimiento que se obtiene es muy elevado siempre que el programa resida en la memoria interna. Si no fuese así, el acceso a instrucciones externas a través de un único bus de 32 bits, como el que posee el TMS320C6701, exigiría 8 accesos para completar la instrucción de 256 bits lo que repercute en el rendimiento. La tabla 24.1 *muestra los recursos de memoria de tres DSPs típicos [24.2].*

	ADSP21160	TMS320C6701	TMS320C6203
Memoria de programa	512 Kbytes	64 Kbytes	384 Kbytes
Palabra de instrucción	48 bits	256 bits	256 bits
Tamaño memoria de programa	85K x 48 bits	2K x 256 bits	12K x 256 bits
Tamaño memoria de datos	Compartida con memoria de programa	64 Kbytes	512 kbytes
Tamaño de caché	32 x 48 bits	2 k x 256 bits	4 k x 256 bits
Palabra de datos	16, 32 ó 64 bits	32 bits	32 bits
Velocidad (picos)	600 MFLOPS	1000 MFLOPS	2400 MIPS

Tabla 24.1. Recursos de memoria interna.

Aritmética de coma fija o coma flotante y longitud de palabra: en la elección de un DSP de coma fija o coma flotante intervienen varios aspectos: coste, consumo de energía, margen dinámico de datos, esfuerzo de desarrollo, etc. En general, un DSP de coma fija es más económico, consume menos y puede integrarse en menos espacio que un DSP de coma flotante, cuestiones muy interesantes en el caso de aplicaciones portátiles o en aplicaciones donde la producción de unidades es muy grande. Por el contrario, el esfuerzo de desarrollo de software en un DSP de coma fija es mayor. Los procesadores en coma fija exigen prestar mucha atención al escalado de forma que se eviten *overflows* o *underflows* durante los cálculos por su reducido margen dinámico. Un DSP de coma flotante posee un margen dinámico mayor que uno de coma fija y es más sencillo de programar (no hay que preocuparse del escalado). Un procesador de 32 bits en coma fija posee un margen dinámico de ≈ 187 dB y 32 bits de precisión, es decir, 4.294.967.296 niveles de cuantificación; un DSP de coma flotante de 32 bits con una mantisa de 24 bits tiene una precisión de 24 bits, 16.777.216 niveles de cuantificación, pero su margen dinámico es de ≈ 1520 dB (los 8 bits del exponente se

[5] DSP de coma flotante de Analog Devices.
[6] DSP de coma fija de Texas Instruments.
[7] DSP de coma flotante Texas Instruments.

utilizan para extender su margen dinámico). En algunos DSPs de coma flotante las operaciones se realizan en un formato extendido de 40 bits alcanzando tanta precisión como uno de coma fija pero con mayor margen dinámico. No obstante, con la optimización adecuada, puede lograrse que un DSP en coma fija trabaje bien en numerosas aplicaciones. Por ejemplo, en aplicaciones donde el margen dinámico de la señal está bien definido un DSP de coma flotante puede aportar poco beneficio.

La longitud de palabra (asociada al ancho de su bus de datos) influirá en la precisión con que se realizan las operaciones; a mayor número de bits, resultados más precisos, aunque el número óptimo dependerá del número de bits de origen. Por ejemplo, el uso de convertidores A/D y D/A de 24 bits requieren un DSP de al menos, 32 bits para asegurar que el nivel de ruido del algoritmo (debido a errores de truncado y redondeo) no exceda el de la señal de entrada (asociado al ruido de cuantificación). El ancho de palabra más común en los DSPs de coma fija es de 16 bits y en coma flotante de 32 bits. Este factor tiene relevancia en el coste, tamaño y capacidad del DSP. Se dispone de procesadores con 16, 24 ó 32 bits en coma fija y generalmente de 32/64 bits en coma flotante. En general, podemos seleccionar un DSP de coma flotante cuando pretendamos desarrollar un prototipo inicial en poco tiempo o cuando el número de unidades previstas no sea excesivo o se necesite un esfuerzo de cálculo intenso sobre datos en coma flotante y con gran margen dinámico de datos. Esto incluye aplicaciones como: procesamiento de imagen, gráficos y aplicaciones militares y espaciales.

Arquitectura del bus externo: este aspecto puede también llegar a ser crítico. La disponibilidad de una interfaz de bus externo es extremadamente útil cuando el código o los datos no caben en la memoria interna del procesador. A veces, el acceso a datos externos puede constituir un cuello de botella más crítico que la propia velocidad computacional del DSP. A menudo, las especificaciones de un DSP se proporcionan considerando que el acceso es siempre a memoria interna y esto no es posible si el programa y/o datos son muy extensos. Cuando se realizan accesos a memoria externa es importante disponer de un bus de datos de suficiente anchura que pueda permitir el acceso en un solo ciclo. El acceso a datos o instrucciones a través de buses de tamaño reducido exige, en general, varios ciclos hasta completar la transferencia lo que puede penalizar la velocidad del algoritmo. En ocasiones, pueden existir dos buses independientes que pueden operar en paralelo para lograr accesos simultáneos. La dimensión del bus de direcciones también estará de acuerdo con el tipo de aplicación y su requerimiento de espacio de memoria. La tabla 24.2 muestra la arquitectura de bus externo de 3 DSPs populares. Por supuesto, debe utilizarse una memoria con tiempos de acceso reducidos para aprovechar las prestaciones, de lo contrario penalizaremos el acceso a memoria externa. En general, para aplicaciones dedicadas, el bus de direcciones puede ser de menor número de bits (24 bits) que en aplicaciones abiertas o de propósito más general que presentan conexión a diferentes tipos de periféricos (32 bits).

Buses externos	ADSP21160	TMS320C6701	TMS320C6203
Número	1	1	2
Reloj	100 MHz	167 MHz	300 MHz
Bus direcciones	32 bits	24 bits	32+32 bits
Bus datos	64 bits	32 bits	32+32 bits
Velocidad de bus	528 Mbytes/s	668 Mbytes/s	1800 Mbytes/s

Tabla 24.2. Arquitectura del bus externo.

Procesador de E/S DMA: el número de canales de DMA disponibles es un factor muy importante y debe considerarse en aplicaciones que requieran, en general, transferencias de datos tanto internas como externas. Algunos canales se asocian a los puertos series, otros se utilizan para la comunicación con dispositivos externos (periféricos), etc. En la tabla 24.3 se indica el número de canales de diversos DSPs.

	ADSP21160	TMS320C6701	TMS320C6203
Número de canales	14	4	4

Tabla 24.3. Canales del controlador de DMA

Puertos de comunicación: la capacidad de comunicación con el mundo externo es otro factor que puede ser crítico. Generalmente, los DSPs proporcionan un puerto serie síncrono de alta velocidad (del orden de 40 MB/s). Debe considerarse el número de puertos disponibles y su velocidad de transferencia. En ocasiones, la lógica interna del puerto serie soporta operaciones típicas en la comunicación serie como la multiplexación por división de tiempo (TDM, *Time Division Multiplexing*). En la tabla 24.4 se indican las características de los puertos serie de comunicación de diversos DSPs.

	ADSP21160	TMS320C6701	TMS320C6203
Número de puertos	2	2	3
Velocidad	50 Mbits/s	83 Mbits/s	150 Mbits/s

Tabla 24.4. Puertos de comunicación serie.

Comunicaciones entre procesadores: esta característica es muy importante en el caso de algoritmos de procesamiento muy intensivos. La disposición de canales dedicados para la comunicación proporciona múltiples ventajas. Por ejemplo, el ADSP21160 presenta unos puertos para la comunicación entre procesadores (*link ports*) que permiten configurar diversas topologías, formando una matriz de una o dos dimensiones de procesadores. La comunicación se establece a alta velocidad (100 MB/s) y es independientemente de otras operaciones externas que pueda realizar el DSP. También permite la conexión con canales DMA y la definición de "*clusters* de procesadores"[8]. Si no existe un mecanismo dedicado de comunicación entre DSPs, por ejemplo TMS320 no dispone de esta capacidad a parte del puerto serie, debe utilizarse hardware externo para implementar las transferencias. En general el uso de multiprocesadores viene de la mano de sistemas que requieran gran cantidad de procesamiento (sistemas de procesamiento sísmico, radares, etc.). La tabla 24.5 muestra las características de varios procesadores en relación con su capacidad de soporte de multiprocesamiento.

	ADSP21160	TMS320C6701	TMS320C6203
Nombre	Link ports	-	-
Número de bits	8 bits	-	-
Cantidad	6	-	-
Velocidad	100 Mbytes/s	-	-

Tabla 24.5. Puertos especiales de comunicación entre procesadores.

Herramientas de desarrollo de software: este factor es esencial para obtener ciclos de desarrollo reducidos. La disponibilidad de editores, compiladores, depuradores y simuladores integrados facilitan el desarrollo. También es interesante disponer de librerías de funciones para aplicaciones mono y multiprocesador, a la vez que funciones matemáticas para el trabajo con vectores y matrices. En aplicaciones de procesamiento intensivo suele ser habitual que se ejecuten varios procesos de forma simultánea. A veces, las herramientas de desarrollo facilitan estas tareas generando pequeños sistemas operativos (*microkernels*) sobre los que

[8] Un *cluster* de procesadores está formado por un grupo de procesadores que comparten recursos y espacio de memoria interna.

se ejecutan las aplicaciones desarrolladas. Básicamente, el sistema operativo actúa como una plataforma de soporte de los procesos programados por el usuario. Este sistema se encarga de gestionar las tareas, asignándoles la prioridad correspondiente y controlando su flujo de ejecución. Liberar al usuario de este trabajo supone un importante ahorro en tiempo de desarrollo al permitirle dedicarse exclusivamente a su aplicación. Otro aspecto muy interesante es la disposición de compiladores que permitan un alto grado de optimización para maximizar el paralelizado de las unidades funcionales y el uso de todos los recursos en general. En arquitectura VLIW, la disposición de un buen compilador es esencial ya que este realiza la planificación de las instrucciones para buscar el máximo paralelismo.

Funcionalidad integrada: A menudo necesitamos disponer de otros recursos en la propia pastilla de DSP como convertidores A/D, D/A o PWM. En general, a más recursos mayor coste, por lo que sólo en los casos en que se utilice, será rentable esta integración. También existe una diferencia clave en el precio según el encapsulado. Los encapsulados PQFP y TQFP son más baratos que los PGAs.

Consumo de energía: como las aplicaciones portátiles son cada vez más habituales, es importante la consideración de DSPs que trabajen con bajas tensiones (3,3 V o 3 V). Como ocurre con los μC, muchos incorporan también mecanismos de ahorro de energía que inhiben el reloj del DSP. En general, se puede afirmar que los DSPs de coma fija consumen menos energía.

24.2 Procesamiento de la información

Filtros Digitales

En el capítulo 4 se introdujeron los principios básicos de los filtros analógicos y se estudiaron las técnicas de diseño más comunes. El filtrado de una señal analógica puede realizarse también por técnicas digitales. El filtrado digital de señal es uno de los algoritmos de uso más habitual y generalmente este tipo de algoritmos se implementan utilizando un computador o un DSP. Este tipo de filtros reemplazan en muchos casos a los clásicos filtros analógicos. Las ventajas que presentan son, básicamente, las siguientes:

a) Un filtro digital es **programable**, es decir, su funcionamiento se determina por un programa almacenado en la memoria del procesador. Esto significa que puede cambiar su comportamiento sin más que modificar el software. En un filtro analógico los cambios se traducen en un rediseño del hardware. Los filtros digitales son diseñados, verificados e implementados fácilmente gracias a la disposición de plataformas avanzadas como computadores.

b) En un filtro analógico sus componentes electrónicos están sujetos a una deriva por motivos diversos como el envejecimiento de componentes o la temperatura. Los filtros digitales no sufren de estos problemas.

c) Los filtros digitales son más versátiles que los analógicos, por ejemplo, pueden adaptarse a cambios en las características de la señal modificando su función de transferencia dinámicamente (filtros adaptativos).

d) Los filtros digitales pueden manejar señales de baja frecuencia cosa que suele ser más difícil de conseguir en filtros analógicos. Las señales de alta frecuencia, dominio antes reservado a filtros analógicos, también pueden ser acometidas por DSPs de alta velocidad.

Los filtros digitales pertenecen a la clase de sistemas de tiempo discreto LTI (*Lineal Time Invariant*, sistemas lineales invariantes con el tiempo) que se caracterizan por las propiedades de causalidad, recursividad

y estabilidad. Se pueden caracterizar en el dominio del tiempo por su respuesta al impulso unitario y en el de la frecuencia por su función de transferencia. La respuesta al impulso unitario de un sistema causal LTI puede ser finita o infinita, y esta propiedad permite su clasificación en: filtros con respuesta al impulso finita o **filtros FIR** (*Finite Impulsive Response*) y filtros con respuesta al impulso infinita o **filtros IIR** (*Infinite Impulsive Response*). El caso más general, un sistema discreto en el tiempo LTI se expresa como:

$$y(kT) = \sum_{p=0}^{m} b_i x(kT - pT) - \sum_{q=1}^{n} a_q y(kT - qT)$$

donde la secuencia de entrada se expresa por x(kT) y la secuencia de salida por y(kT). Si algún coeficiente a_q es distinto de cero, el sistema correspondiente es recursivo y su respuesta al impulso es de duración infinita (IIR). Si todos los coeficientes a_q fuesen cero el sistema correspondiente sería no recursivo y su respuesta al impulso sería de duración finita (FIR). Para una discusión detallada puede recurrir a las referencia que se mencionan en la bibliografía [24.4-24.5].

El diseño de filtros digitales consiste en los siguientes pasos:

1. Determinar la respuesta deseada. La respuesta se especifica generalmente en el dominio de la frecuencia en términos de magnitud y de la fase.
2. Selección de la clase de filtro (FIR o IIR).
3. Seleccionar el mejor miembro dentro de la clase de filtro seleccionada.
4. Implementar el filtro en un computador o DSP.
5. Analizar su comportamiento para ver si satisface todos los requerimientos.

Existen herramientas software que nos permiten realizar el diseño y el análisis de un filtro digital de forma muy sencilla[9]. Básicamente, nos limitamos a especificar de forma numérica o incluso gráfica las características del filtro, y el programa obtiene los coeficientes (b_p y a_q) que deben utilizarse.

24.3 Comunicación hombre-máquina

Un sistema de medida requiere, en última instancia, comunicar la información. En este capítulo se estudian dispositivos dedicados a comunicar información a un operador humano.

24.3.1 PRESENTACIÓN DE LA INFORMACIÓN

En todo sistema de presentación de información es deseable que esta se realice siempre de la forma más clara y simple posible. La figura siguiente resume los dispositivos más utilizados para representar información atendiendo a su tipo y tecnología. Una de las primeras decisiones que se toman a la hora de elegir un dispositivo de presentación es la necesidad o no de registrar la información que se obtiene del sistema de medida.

[9] En el CD se proporciona una herramienta de diseño de filtros digitales.

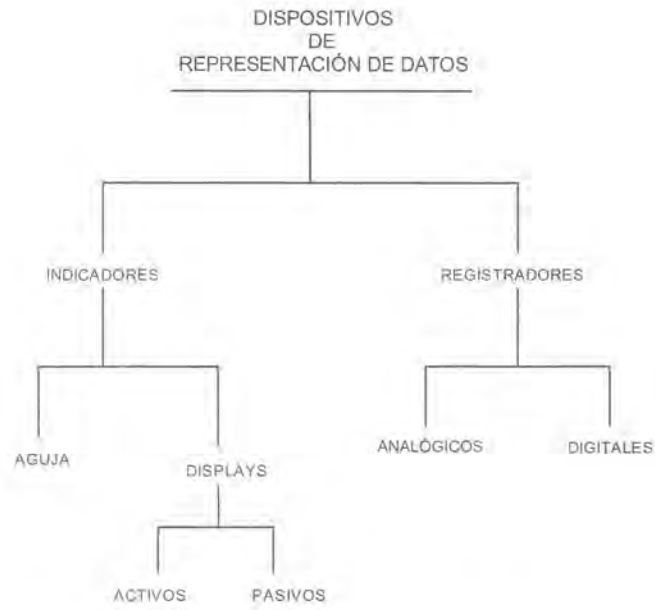


Figura 24.13. Clasificación de los dispositivos de representación de datos.

Dispositivos registradores

El registro de variables de procesos es necesario siempre que sea necesario realizar estudios posteriores sobre ellas por los motivos que sean. El registro se utiliza en situaciones especiales donde la señal de información cambie a una velocidad que no puede ser visualizada convenientemente por un operador humano, por ejemplo, es necesario registrar la señal de un electrocardiograma para ser interpretada posteriormente. En otras ocasiones el alto volumen de información generada es lo que obliga a su registro, por ejemplo, en la medida de los ciclos de temperatura utilizados en un horno de cocción para verificar que se cumplen los intervalos programados. Los sistemas de registro pueden ser analógicos o digitales. En los registradores analógicos se dispone de un sistema móvil como una bobina que se sitúa entre un campo magnético radial, de forma que la corriente que circula, variable analógica, produce su deflexión. Solidariamente con la bobina se une una plumilla que permite dibujar directamente sobre un papel las variaciones de la variable que pretendemos medir (electrocardiogramas). En la figura 24.14 se muestra un sistema parecido donde se sustituye la plumilla por un haz óptico que imprime sobre una superficie fotosensible. En los sistemas de registro

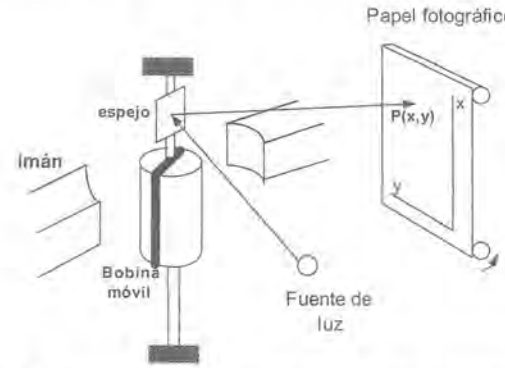


Figura 24.14. Galvanómetro de UV (registrador analógico) [24.1].

digital, la señal de información, generalmente analógica, se convierte primero al dominio digital y se transfiere con posterioridad al registrador como por ejemplo una impresora.

Dispositivos indicadores

En aquellas aplicaciones donde sea suficiente una indicación y no el registro, la elección suele decantarse entre dispositivos de aguja y displays indicadores.

a) *Dispositivos de aguja*

Estos sistemas están basados también en una bobina móvil que se sitúa dentro de un campo magnético. La bobina está sujeta a un eje a través de un pequeño muelle cuya fuerza de tracción se vence al circular la corriente. En este caso la plumilla se sustituye por una aguja que se mueve sobre una escala graduada. Presentan como ventaja que pueden conectarse directamente al sistema de fuente analógica. En un dispositivo de aguja el operador debe interpolar la lectura si la aguja se sitúa entre dos puntos de la escala. Como la interpolación es subjetiva se produce un error de observación. Para reducir los errores de observación es importante que la escala sea legible, clara y sin excesivas marcas de escala.

Los dispositivos de aguja están en desuso para medidas de precisión pero tienen gran ventaja para lectores rápidos o detectar valores anómalos. Por ejemplo, para un observador que tenga que supervisar un gran número de variables representadas en un panel le resulta fácil detectar el movimiento de una aguja pero sería más complejo detectar niveles fuera del margen sobre un *display* numérico.

b) *Dispositivos de presentación o displays*

En los *displays* se elimina la necesidad de interpretación ya que se representa directamente un número (o una gráfica). Cuando el número de variables a representar sea reducido puede utilizarse un dispositivo alfanumérico de pequeña escala como un LCD. Para sistemas que requieran la visualización de muchas variables se pueden utilizar display del tipo CRT o pantallas de mayor dimensión. Por su importancia se expone a continuación brevemente las características de algunos de los displays más utilizados. Los dispositivos de *presentación o displays se dividen en dos grupos:*

- Activos o emisivos.
- Pasivos o no emisivos.

En los *displays* activos los píxeles (puntos que componen la imagen) de la pantalla emiten radiación visible, mientras que en los pasivos se limitan a dejar pasar o no la luz procedente de una fuente externa, que puede ser la propia luz ambiente (por reflexión o transmisión) o una fuente de luz incorporada al dispositivo (generalmente por transmisión). Entre los visualizadores activos el más utilizado es el CRT o tubo de rayos catódicos junto a los LEDs y, entre los pasivos, los de mayor uso son los basados en cristales líquidos (LCD).

CRT (*Cathode Ray Tube*) tubo de rayos catódicos

La figura 24.15 muestra el esquema básico de un tubo de rayos catódicos. En estos dispositivos la emisión luminosa se consigue excitando un material fosforescente con el que se recubre la pantalla por medio de un haz de electrones procedentes de un cátodo. La intensidad del haz electrónico se modula por medio de una rejilla situada entre el cátodo y el ánodo. El haz pasa después a través de un sistema de enfoque (electrónico) y atraviesa las placas del sistema de deflexión horizontal (X) y vertical (Y) que lo desplazan en la dirección correspondiente hasta llegar a la pantalla recubierta por la sustancia fosforescente. El fósforo emite luz cuando incide sobre él la fuente de electrones, apareciendo un punto luminoso en la pantalla. Para formar una imagen se aplican señales en forma de diente de sierra a las placas de deflexión horizontal y vertical de forma que el haz electrónico se desplaza de izquierda a derecha y de arriba a bajo, barriendo la pantalla.

Modulando de forma adecuada el haz, a través de la rejilla de modulación, y gracias a la persistencia del fósforo (la luz emitida por el fósforo no cae de forma instantánea cuando deja de estar excitado) puede componerse una imagen. Para que esta imagen permanezca estable y no exista parpadeo deben de realizarse los barridos o exploraciones de la pantalla a la frecuencia adecuada. Esta tecnología es muy madura y tiene como principal ventaja su coste y su capacidad de resolución.

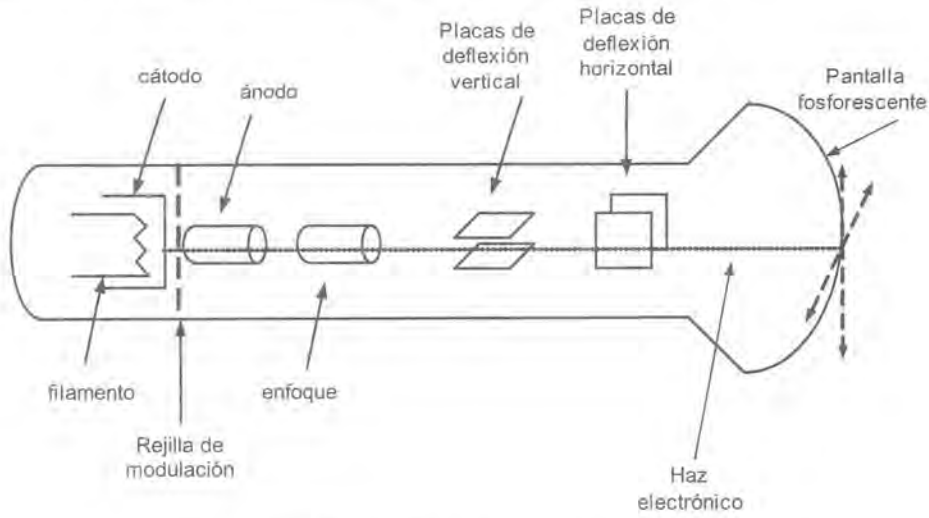


Figura 24.15. Tubo de rayos catódicos.

LCD (*Liquid Cristal Displays*)

El cristal líquido es un material especial que retiene las propiedades de una estructura cristalina, presenta una cierta ordenación de sus moléculas, pero en fase líquida. Esta ordenación puede cambiar si se aplican campos eléctricos, modificando como consecuencia sus propiedades ópticas. La estructura de un LCD consta de dos polarizadores cruzados (uno en la dirección X y otro en Y) y dos electrodos entre los que se coloca la estructura de cristal líquido (figura 24.16). En la parte trasera puede situarse un espejo, si se desea trabajar en la modalidad de reflexión.

Si no se aplica tensión, la radiación óptica atraviesa la muestra pero gira su plano de polarización 90°, siguiendo el eje óptico de las moléculas del cristal (figura 24.16a). Situando los polarizadores con un desfase de 90°, la luz recorre el camino hacia el espejo ya que atraviesa el primer polarizador y no queda bloqueada por el segundo pues giró 90° en el camino. De esta forma, se produce una reflexión en el espejo y la luz recorre una trayectoria de vuelta análoga a la incidente apareciendo el visualizador como transparente. Un razonamiento idéntico se aplica al caso de un LCD por transmisión, aunque, en este caso, se pasa sólo una vez por el material y se prescinde del espejo. Cuando se aplica un campo entre los electrodos de la célula, la orientación del eje óptico de las moléculas es paralela al campo y, por tanto, la luz que atraviesa la muestra no sufre cambio de polarización (figura 24.16b). Ahora la radiación será absorbida por el segundo polarizador y no existirá radiación reflejada (o transmitida). El visualizador aparece en este caso como opaco. Debido a las pérdidas introducidas por los polarizadores, el estado claro es de color gris, especialmente por reflexión, mientras que el oscuro es negro, equivalente a dos polarizadores cruzados. De esta manera, pueden formarse la silueta de números o puntos de gráficas.

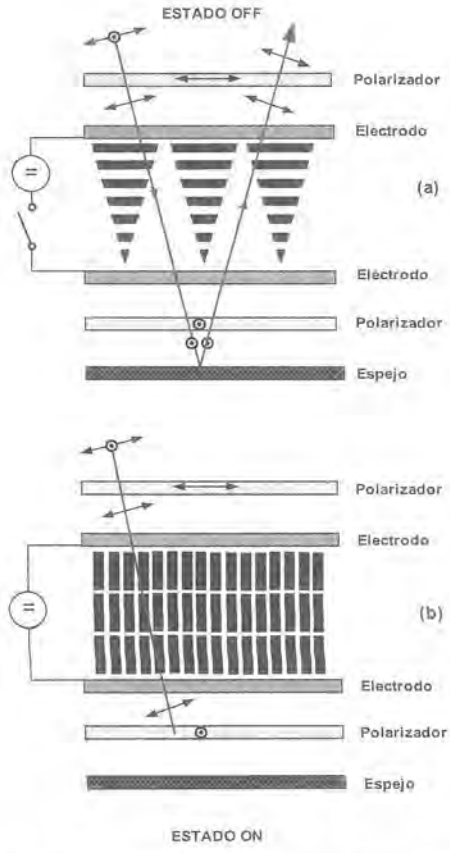


Figura 24.16. *Estructura de un cristal líquido [24.7].*

LED

Los LEDs son muy utilizados como componentes individuales con bajos niveles de integración, por ejemplo, en los *displays* de 7 segmentos y en las matrices de puntos. Este tipo de dispositivos permiten la representación de letras y números de forma compacta. La figura 24.17 describe su estructura interna.

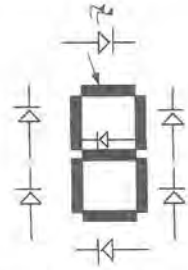


Figura 24.17. Estructura de un *display* de 7 segmentos.

En el caso de los *displays* de 7 segmentos están formados por un conjunto de LEDs dispuestos en forma de ocho (figura 24.17). Los números se representan polarizando en directa los LEDs asociados al contorno

del número que pretendemos representar. Al polarizarse en directa se ilumina el diodo correspondiente mientras el resto permanecen apagados. Los *displays* se encapsulan adoptando la forma de un IC. En este empaquetado y con el objeto de ahorrar patillas se unen todos los ánodos (ánodo común) o todos los cátodos (cátodo común) a un punto común y se controla el encendido de los LEDs con la señal correspondiente en el cátodo o en el ánodo. Estos dispositivos suelen controlarse mediante decodificadores en cuya entrada se sitúa un código digital en formato binario o BCD, y en su salida activan las líneas correspondientes para encender los LEDs que representan el número decimal equivalente. En la figura 24.18 se muestra un ejemplo típico de conexión de un *display* de 7 segmentos y su decodificador asociado. Las matrices de puntos sustituyen los LEDs en forma de segmentos por una matriz de puntos LEDs de forma circular. Los caracteres y números se forman igual, resaltando la silueta al iluminarse los LEDs apropiados con cada carácter.

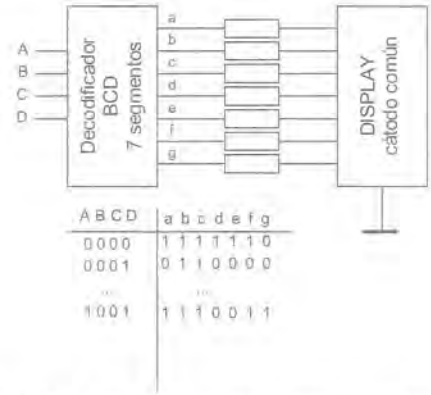


Figura 24.18. *Display* de 7 segmentos y decodificador.

Además de los elementos señalados, existen otras tecnologías para displays como pantallas electroluminiscentes (EL), pantallas de plasma, pantallas VFD (fluorescencia en vacío) y cristal líquido en matriz activa (TFT) [24.7]. Las pantallas electroluminiscentes (figura 24.19) se basan en la capacidad que presentan ciertos materiales para emitir luz de forma espontánea cuando se les aplica una tensión eléctrica continua o alterna[10]. Estos dispositivos se construyen formando filas y columnas de electrodos entre los que se sitúa el material

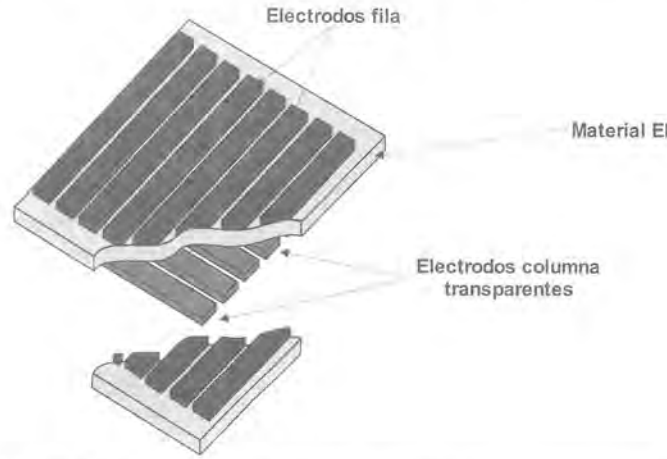


Figura 24.19. Estructura de una pantalla electroluminiscente.

[10] Este fenómeno se denomina de electroluminiscencia (EL) y su descubrimiento es bastante antiguo (H.J. Round, 1907).

electroluminiscente[11]. El conjunto se recubre de una lámina de vidrio. La pantalla muestra píxeles en color amarillo brillante sobre fondo negro. Presenta una gran luminosidad y ángulo de visión superior a un CRT.

Las pantallas de plasma presentan un aspecto similar al mostrado en las electroluminiscentes con la diferencia de que se sustituye el material electroluminiscente por un gas a baja presión (neón generalmente). La matriz de filas y columnas se recubre de un dieléctrico y se aplica una tensión algo inferior a la de ruptura del gas. Sobre esta tensión se superponen los pulsos de datos que llevan a la ionización del gas produciendo el plasma y emitiendo una radiación característica, generalmente naranja.

Las pantallas de fluorescencia de vacío presentan una tecnología similar a la de los CRT. Para conseguir la emisión luminosa se excita un material fluorescente con un haz de electrones procedente de la emisión termoiónica de un filamento. La modulación del haz se controla con una rejilla (figura 24.20).

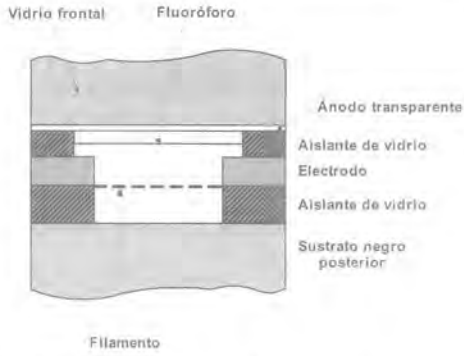


Figura 24.20. Pantalla de fluorescencia de vacío.

Produce generalmente un color verde azulado muy brillante y son muy utilizadas en vídeos domésticos porque se ven muy bien a cierta distancia.

	Tipo de display	Ventajas	Limitaciones
	LED	Matrices pequeñas, buena luminosidad.	Coste de ensamblado en matrices grandes.
ACTIVOS	Electroluminiscencia	Buen ángulo de visión, alta calidad estética.	Coste.
	Plasma	Diversidad de tamaños y aplicaciones, alto brillo y muy robusto.	Baja eficiencia, elevado tamaño, peso y coste.
	CRT	Muy bajo coste, gran resolución y posibilidad de color.	Volumen, peso y profundidad elevados. Presentan problemas con alta luminosidad.
	Fluorescencia	Muchos tamaños, matriz direccionable.	Complejidad estructural elevada.
PASIVOS	Cristal líquido (TWIST)	Potencia mínima, poco peso, bajo coste, excitación con luz ambiente.	Pequeño ángulo de visión, bajo contraste.
	Cristal líquido en matriz activa (TFT)	Permiten color.	Necesita iluminación posterior.

Tabla 24.6. Comparativa de *displays* de representación de datos [24.7].

[11] Generalmente se utiliza sulfuro de zinc impurificado con manganeso.

Las pantallas TFT son en la actualidad muy utilizadas en computadores y están sustituyendo paulatinamente al CRT como *display*. En este tipo de pantallas, los electrodos que constituyen las filas y columnas se sustituyen por elementos activos (transistores de capa delgada, TFT).

24.4 Sensores Inteligentes

Acercar cada vez más la inteligencia al punto de medida, evitando así la transmisión de señales analógicas, fue y sigue siendo el objetivo de los sistemas de medida industriales que, a pasos de gigante, se desarrollan en arquitecturas distribuidas o de buses que han sido referidas en el capítulo 1.

Un sensor puede considerarse inteligente (*smart*) si, además del sensor o sensores propiamente dichos, en el mismo circuito integrado se incorpora otro u otros circuitos que realizan alguna(s) de las siguientes funciones:

- Acondicionamiento analógico.
- Conversión analógica-digital.
- Procesamiento primario de la información (validación, linealización, etc.).
- Conversión a unidades de ingeniería.
- Auto identificación.
- Auto diagnóstico.
- Salida de datos digital en unidades de ingeniería.
- Comunicación con una red digital.

La figura 24.21 presenta la arquitectura general de un sensor inteligente que, en este caso, realiza todas las funciones citadas anteriormente.

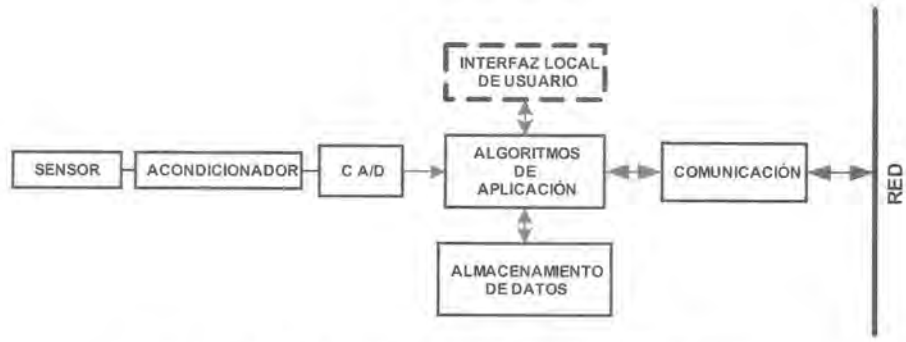


Figura 24.21. Esquema general de un sensor inteligente.

El primer sensor "*smart*" de presión fue fabricado por Honeywell en 1969 para el sistema de aire de los aviones DC-9, que integraba dos piezorresistencias (elementos sensores), dos condensadores y un inversor (figura 24.22) conformando un oscilador cuya frecuencia era proporcional a la presión. Este sistema puede considerarse de una primera generación en la que el acondicionamiento y el sensor se encuentran en una misma pastilla.

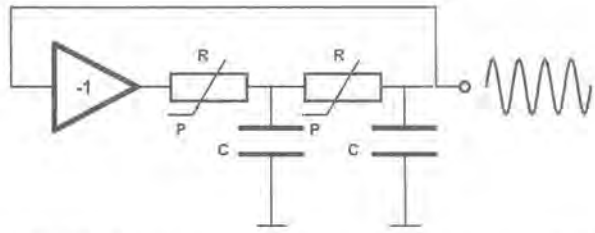


Figura 24.22. Sensor *"smart"* de presión de Honeywell (1969).

Ya en la década de los años ochenta, con el desarrollo de los microprocesadores para el procesamiento digital de señales, varias compañías desarrollaron circuitos integrados (con multiplexor para varios sensores), convertidor A/D, procesador digital de señales y conversión D/A, con salida analógica 4...20 mA, con la arquitectura que se presenta en la figura 24.23, que constituyeron los primeros transmisores inteligentes; aunque no se corresponden con el concepto de sensor inteligente expuesto anteriormente, a la luz del estándar IEEE-P1451, se amplía el concepto de sensor *"smart"*. Según esta norma cualquier transductor[12] que realice funciones mas allá de las necesarias para generar una representación correcta de la cantidad medida, que simplifiquen la integración del transductor a las aplicaciones en un contexto conectado en red, es un transductor inteligente, por lo que la unión transmisor inteligente + sensor es considerada por algunos como un sensor *"smart"*.

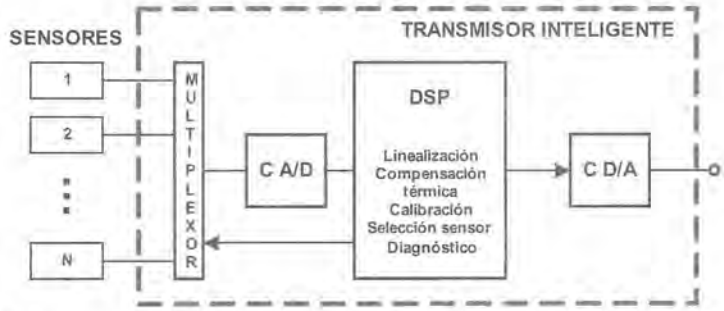


Figura 24.23. Arquitectura de los transmisores *"smart"* de la década de los ochenta.

Específicamente Honeywell en 1983 desarrolló el sistema industrial ST3000 para el control de procesos en el que los sensores eran de presión diferencial estática y de temperatura y el convertidor A/D era un convertidor V/f. El software incorporaba funciones de compensación de la presión estática, calibración de margen remota, diagnóstico y direccionamiento de sensores y la salida del convertidor D/A en el estándar de 4...20 mA. Ya en 1989 Honeywell fabricó los primeros transmisores inteligentes totalmente digitales.

Por esta razón se han desarrollado circuitos integrados conocidos con el nombre de microconvertidores que incluyen la comunicación con la red, todo el procesamiento digital, la conversión A/D y parte o todo el acondicionamiento analógico, cuya razón fiabilidad/coste hoy día es muy elevada.

Un ejemplo de estos sistemas es el microconvertidor ADuC816, fabricado por Analog Devices, que satisface los requerimentos del estándar IEEE-P1451.2 y cuyo diagrama de bloques funcionales simplificado se presenta en la figura 24.24.

---

[12] El **sensor** es un transductor que convierte la energía del proceso en otra adecuada (generalmente eléctrica) para poder ser procesada por un sistema de medida. Al propio tiempo el **actuador** es un transductor que convierte la energía procedente de un sistema de medición y control (generalmente señal eléctrica) en otra encargada de modificar el proceso.

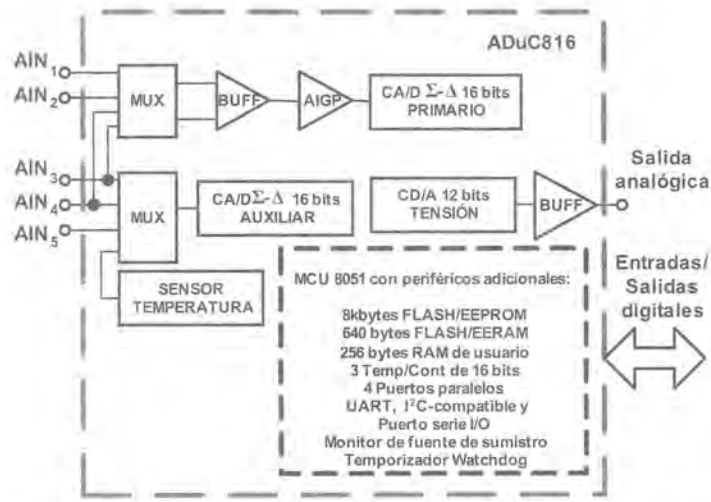


Figura 24.24. Diagrama de bloques funcionales simplificado del microconvertidor ADuC816.

Este sistema posee dos convertidores A/D de alta resolución, un microcontrolador de 8 bits con 26 líneas de entrada/salida programables, trabaja con una sola fuente de 5 ó 3 V y su consumo con esta última tensión es de sólo 10 mW. Incluye una salida analógica, un sensor de temperatura y un amplificador de ganancia programable que le permite trabajar con señales de entrada muy pequeñas (20 mV) ofreciendo resoluciones correspondientes a una calidad binaria real de 13 bits.

Las tecnologías de fabricación de circuitos integrados que permiten la fabricación de diferentes tipos de sensores en el propio silicio con el resto de los circuitos con alta fiabilidad, bajo consumo, moderada precisión y muy bajo coste, han dado lugar al desarrollo de los conocidos como sensores integrados o MEMS (*Micro ElectroMechanical Systems*).

En estos sistemas se desarrolla, monolíticamente, el sensor mediante técnicas de micromecanizado junto a los circuitos de acondicionamiento analógico y, en los más desarrollados, el microcontrolador con conversión A/D, D/A y comunicación con un bus digital o con una red.

La figura 24.25 presenta, a modo de ejemplo, un diagrama de la microfotografía de un MEMS fabricado con tecnología CMOS de aproximadamente 9 mm^2 [24.9].

Se trata de un sensor "*smart*" que mide la presión diferencial mediante sensores capacitivos, posee un circuito de acondicionamiento con capacidad de programación y un circuito de compensación de temperatura. Se mide la capacidad de los elementos sensores y los de referencia y la señal de presión se obtiene por la razón entre estas capacidades y es convertida en una tensión analógica ratiométrica de salida. Los tres terminales superiores son los que normalmente se utilizan y los de la derecha son para programación de, entre otras cosas, los parámetros de ganancia de los amplificadores y de la red de realimentación del convertidor capacidad-tensión (C/V), en función del diámetro de las celdas sensoras micromecanizadas para diferentes márgenes de la presión de entrada.

Por sólo citar algunos ejemplos, los MEMS se utilizan en la industria automovilística, tanto en la robótica que los ensambla como en los sistemas de los frenos, encendido automático de luces y limpiaparabrisas, *airbag*, caja de cambios y en el motor de los coches; en nuestras casas se encuentran presentes la calefacción, ventilación y aire acondicionado, para la regulación de la presión, humedad y temperatura; en medicina, en los monitores de presión sanguínea, corazón artificial, etc.

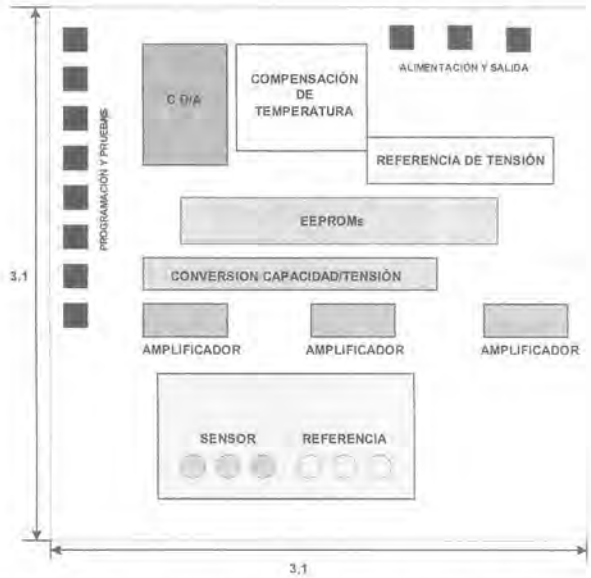


Figura 24.25. Estructura interna de un MEMS capacitivo para presión diferencial con salida analógica.

Ejemplos de circuitos comerciales pueden ser los acelerómetros de Analog Devices ADXL150-250, de uno y dos ejes respectivamente, en los que el sensor está micromecanizado en el propio silicio donde se realizan el resto de los componentes del circuito integrado que ofrece una salida de tensión proporcional a la aceleración. Las hojas de características del fabricante se ofrecen en el CD.

La necesidad de una norma

Las compañías productoras de sensores y sistemas de medida se encontraron ante la necesidad de lograr ciertos estándares en lo referente a la compatibilidad de los sistemas, dada la gran variedad de redes y buses digitales existentes.

Y así surge, en 1997, el estándar IEEE-P1451 (*Transducer to Microprocessor Interface*) que regula las directivas de los sensores inteligentes, así como la comunicación con el bus digital [24.10]. Este estándar se ha convertido en el más aceptado por la mayoría de las compañías y utiliza el término **transductor inteligente** como sistema conectado a una red común, con la información necesaria para saber si es un sensor o un actuador. Esta norma, mediante su estándar 1451.2 (*Transducer to Microprocessor Communication Protocol and TEDS Formats*), divide el esquema general del sensor inteligente de la figura 24.23 de la forma presentada en la figura 24.26.

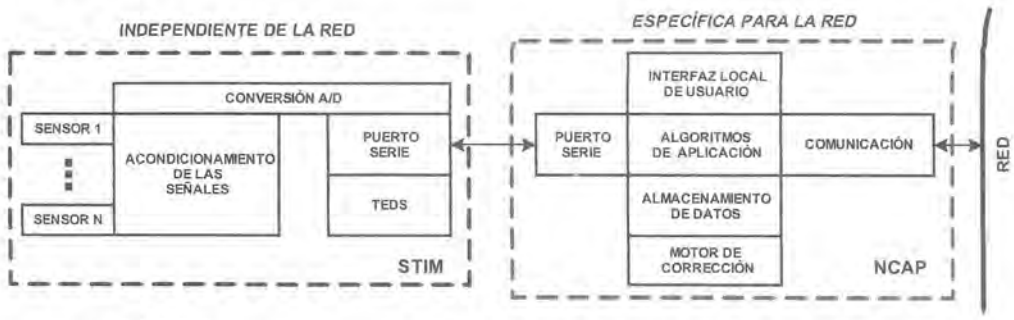


Figura 24.26. Esquema general del transductor inteligente de la norma IEEE 1451.2.

El sistema de la izquierda denominado STIM (*Smart Tranducer Interface Module*) contiene el sensor o sensores, el acondicionamiento, la conversión A/D, un puerto serie que intercambia información digital con el NCAP (*Network Capable Application Processor*) mediante una conexión punto a punto de cuatro hilos: dos de comunicación (DIN, DOUT) y dos de alimentación (+5 V y Común) y una memoria TEDS (*Transducer Electronic Data Sheet*) que almacena toda la información sobre el(los) sensor(es), en forma de tabla organizada por aspectos tales como tipo, datos del fabricante, parámetros que mide, margen estático y dinámico de la entrada y la salida, precisión, datos de calibración, datos de conversión a unidades de ingeniería, posición en el proceso, etc. La norma **precisa** la comunicación entre el STIM y el NCAP, así como los contenidos de las TEDS, de manera tal que este bloque es independiente de la red.

El sistema de la derecha denominado NCAP es un microprocesador que administra la comunicación con la red, con una arquitectura básica de la figura 24.25. El estándar 1451.1 de 1999 se refiere a las características de la información que se intercambia con la red por parte del NCAP.

El funcionamiento de ambos bloques como sistema es muy sencillo: el STIM entrega al NCAP la información sobre los contenidos de las tablas TEDS y los valores de las variables analógicas obtenidos por los sensores, acondicionados y digitalizados según la organización del proceso de muestreo que haya sido indicada por el NCAP. El NCAP normaliza esta información y la coloca en la red. También transforma la información proveniente de esta al lenguaje propio del STIM con la finalidad de poder corregir algún parámetro del sensor, cambiar la identificación, etc.

Resumen

- En este capítulo se estudian diversos dispositivos digitales programables: microprocesadores, microcontroladores y DSPs. Los microprocesadores son circuitos programables integrados en una única pastilla de silicio que realizan las funciones encomendadas a un **procesador** digital y constituyen el núcleo operacional de un sistema más general: el computador. La característica básica de estos circuitos es su carácter de propósito general.
- Los **microntroladores** son dispositivos que incluyen en una misma pastilla de silicio gran parte de la funcionalidad que les permite actuar como sistemas computadores autónomos adaptados a tareas de control y con una capacidad especial para conectarse directamente con el proceso.
- Los **DSPs** son procesadores especializados en el cálculo numérico a alta velocidad. En general, implementan algoritmos de procesamiento de señal como el filtrado digital o el análisis espectral. Para cada uno de ellos se estudian sus características básicas y se fijan los criterios de selección más importantes.
- El capítulo continúa con el estudio de diversos dispositivos de representación de datos. En general estos dispositivos pueden clasificarse según la necesidad o no de registrar la información que se obtiene del sistema de medida.
- Los sistemas de registro pueden ser analógicos o digitales. En los registradores analógicos un sistema móvil conectado a la fuente analógica se desplaza de forma análoga a la variable física que queremos registrar. En los sistemas de registro digital, la señal de información se convierte primero al dominio digital y se transfiere con posterioridad al registrador como en una impresora.
- En aquellas aplicaciones donde sea suficiente una indicación y no el registro, la elección suele decantarse entre dispositivos de aguja y *displays* indicadores. Los dispositivos de presentación o displays permiten presentar mayor información y de forma más precisa. Los *displays* se dividen en dos grupos: activos o emisivos y pasivos o no emisivos. Entre los activos el más utilizado es el CRT que permite representar

grandes volúmenes de datos y los LEDs (*displays* de segmentos y matrices de puntos). Entre los pasivos los más utilizados son las pantallas de cristal líquido que permiten producir displays de bajo coste y bajo consumo. También se indican otras tecnologías de visualización como: paneles electroluminiscentes, paneles de plasma y paneles de fluorescencia en vacío.

- Finalmente se analizan los sensores inteligentes. Un sensor puede considerarse inteligente (*smart*) si, además del sensor o sensores propiamente dichos, en el mismo circuito integrado se incorpora otro u otros circuitos que realizan alguna(s) de las siguientes funciones: acondicionamiento analógico, conversión analógica-digital, procesamiento primario de la información (validación, linealización, etc.), conversión a unidades de ingeniería, auto identificación, auto diagnóstico, salida de datos digital en unidades de ingeniería o comunicación con una red digital.

25

Sistemas de adquisición de datos

25.1 Introducción a los sistemas de adquisición de datos (SAD)
25.2 Configuraciones de un sistema de adquisición de datos
25.3 Buses de instrumentos

25.1 Introducción a los sistemas de adquisición de datos (SAD)

En cualquier actividad relacionada con la ciencia y la tecnología surge la necesidad de medir variables físicas. Un **sistema de adquisición de datos (SAD)** es el instrumento del que nos servimos para obtener información de un determinado proceso o del estado de una planta. En la figura 25.1 se muestra el diagrama de bloques de un sistema de adquisición de datos. En general, los sistemas de adquisición de datos constan de cinco elementos básicos[1]:

1. Los sensores que convierten un fenómeno físico en una magnitud eléctrica.
2. Un bloque acondicionador que realiza el acondicionamiento para aislar, filtrar, convertir y/o amplificar la señal.
3. Un sistema de adquisición que convierte la señal del dominio analógico al digital.
4. Un sistema de tratamiento que realiza operaciones con los datos (información digital) con objeto de transformarlos en información útil.
5. *Un sistema de representación cuyo objeto es comunicar los resultados al usuario.*

En capítulos anteriores del texto se han estudiado en detalle cada uno de estos bloques. Ahora, nuestro interés es el estudio de sistemas que realizan todas estas funciones de forma integrada.

[1] En ocasiones un sistema de adquisición de datos también puede incluir un bloque denominado "actuador" que permite modificar el proceso si fuese necesario.

Figura 25.1. Diagrama de bloques de un sistema de adquisición de datos.

25.2 Configuraciones de un sistema de adquisición de datos

Los sistemas de adquisición de datos suelen configurarse de acuerdo a una de estas tres posibilidades o una combinación de ellas:

1. **Sistemas basados en un computador** con una o varias **tarjetas de adquisición de datos** conectadas al bus interno del computador y que realizan básicamente la conversión de la señal entre el dominio analógico y el digital.

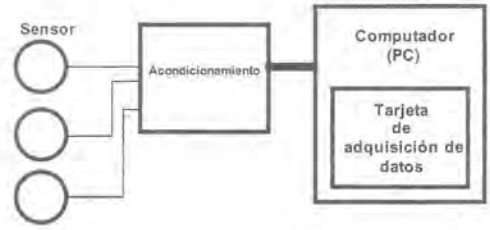


Figura 25.2. Sistema de adquisición integrado en un computador y basado en tarjetas de adquisición de datos.

2. Sistemas basados en **instrumentos de adquisición**, que a su vez pueden ser:

   2.1. Instrumentos autónomos e independientes del bus de un computador pero con capacidad de conectarse a este a través de buses de instrumentación (GPIB) o a través de puertos de comunicación serie (RS-232, USB, etc.) o paralelo.

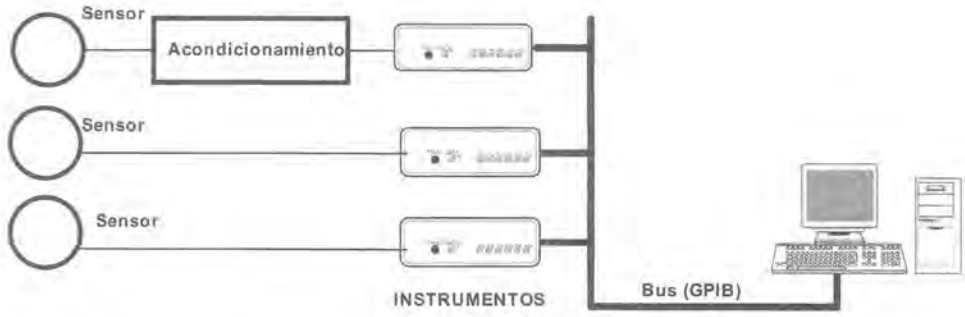


Figura 25.3. Sistema de adquisición basado en instrumentos de adquisición independientes.

2.2. Instrumentos basados en módulos de adquisición (instrumentación modular) conectados a un bus local especializado como VXI (basado en el bus VME) o PXI (basado en el bus PCI).

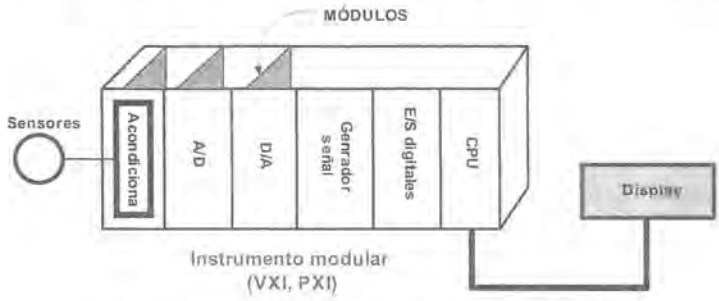


Figura 25.4. Sistema modular de adquisición.

Los sistemas de adquisición de datos basados en un computador se han convertido en una de las alternativas más prometedoras debido a su coste y versatilidad. En esta configuración, suele utilizarse un computador personal (PC) como plataforma para realizar el procesamiento, el almacenamiento y la visualización, y una tarjeta de adquisición de datos para la captura de las señales analógicas. En general, suele utilizarse una alternativa de este tipo cuando no se requieran altas prestaciones de velocidad y el número de señales de captura no sea excesivo.

Aunque el computador se ha convertido en un excelente recurso para construir un sistema de adquisición de datos, esta plataforma no está suficientemente acondicionada para trabajar en un entorno industrial exigente; así, su inmunidad a interferencias es bastante pobre, su refrigeración deficiente, sus slots de expansión muy limitados, etc. Goza, sin embargo, de un alto grado de conectividad que permiten realizar la administración y monitorización remota a través de redes (LAN, Internet...) y sobre todo, una capacidad de visualización y cálculo bastante avanzadas, junto a un soporte de herramientas de programación muy potente.

Los sistemas de adquisición de datos basados en instrumentos independientes suelen utilizarse en aplicaciones más exigentes que requieran capturar un gran número de variables con gran precisión y fiabilidad. En general, son soluciones de mayor coste pero están mejor acondicionadas y ofrecen mayores prestaciones que la anterior. En el mercado existen diferentes soluciones para resolver las necesidades más exigentes: sistemas de adquisición, sistemas actuadores, *data-loggers*, etc. Estos equipos poseen una funcionalidad específica que puede ser administrada por el usuario a través del panel frontal del instrumento, con la típica botonera, o por software. En este último caso, el instrumento dispone de algún tipo de interfaz (bus GPIB o puerto serie) para su conexión a un computador. El instrumento dispone de un conjunto de comandos que son enviados por el computador y que permiten automatizar su funcionamiento. Además, pueden aprovechar el computador como dispositivo de cálculo y visualización de los datos adquiridos. La capacidad de comunicación de estos instrumentos se amplía cada vez más, dotándolos en ocasiones de conexión a redes (Ethernet) para facilitar su gestión remota.

Los sistemas basados en instrumentos modulares permiten configurar sistemas de adquisición más potentes y a un coste que puede llegar a ser más reducido que con instrumentos autónomos, sobre todo en sistemas de cierta dimensión o con necesidad de crecimiento futuras. Los sistemas basados en módulos son muy robustos, flexibles y potentes. Se dispone de módulos de gran cantidad de fabricantes siguiendo estándares muy aceptados en la industria. A continuación, analizaremos en detalle cada una de estas soluciones, describiendo las capacidades que aportan y los criterios de selección más relevantes.

25.2.1 TARJETAS DE ADQUISICIÓN DE DATOS

Las tarjetas de adquisición de datos son dispositivos en formato de tarjeta de circuito impreso que se conectan directamente al computador a través de las ranuras de expansión. Generalmente son tarjetas que se co-

nectan al bus PCI[2], por ser el bus de comunicación con dispositivos periféricos más extendido. No obstante, puede haber tarjetas para otro tipo de computadores y para otros buses.

La tarjeta de adquisición más sencilla consiste en un hardware de adquisición que tan sólo convierte la señal analógica de entrada en una señal digital que es enviada al computador sin realizar ningún tipo de procesamiento. En otras tarjetas puede realizarse un tratamiento de los datos en la propia tarjeta incorporando para ello un procesador especializado, por ejemplo uno o varios DSPs. Estas tarjetas con procesador incorporado suelen utilizarse en aplicaciones exigentes (alta velocidad de adquisición y gran cantidad de procesamiento) donde se demande procesamiento local, bien porque el procesador central del computador no posea la suficiente potencia para procesar las muestras en tiempo real, o bien porque existan limitaciones derivadas de la velocidad máxima de transferencia de datos a través del bus.

Algunas tarjetas son en sí mismas instrumentos y se denominan **instrumentos en tarjeta**. Estas tarjetas aprovechan básicamente la capacidad de comunicación y representación de datos del computador, pero contienen todas las características que las hacen funcionar como un instrumento autónomo. Por ejemplo, existen soluciones como osciloscopios en formato de tarjeta que utilizan el computador como interfaz de visualización del panel de mandos y pantalla, pero la tarjeta incluye toda la funcionalidad del instrumento. Las tarjetas de adquisición no responden a este concepto y tienen un carácter más general. La figura 25.5 muestra el diagrama de bloques de una tarjeta de adquisición típica.

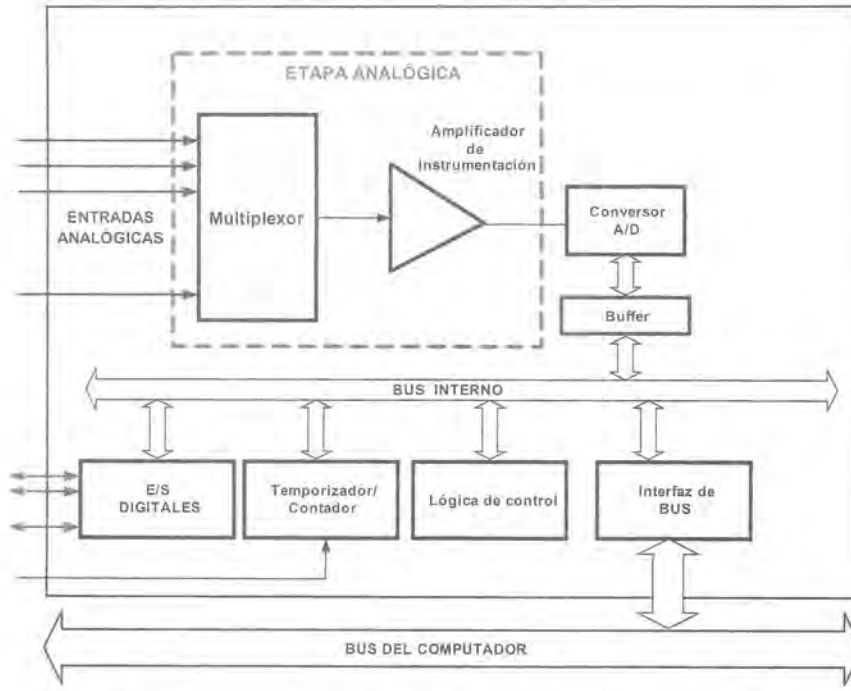


Figura 25.5. Estructura general de una tarjeta de adquisición de datos.

El núcleo central de una tarjeta de adquisición está compuesto por los circuitos analógicos de entrada y el conversor A/D. Los circuitos analógicos de entrada incluyen un multiplexor, al que llegan las distintas variables de entrada, un amplificador de ganancia programable y un circuito de muestreo y retención (S&H);

[2] *Peripherical Conection Interface*, interfaz para la conexión de periféricos. Bus de extensión típico de los computadores tipo PC.

a continuación, se encuentra el propio convertidor A/D. En general, las tarjetas disponen de una memoria temporal o *buffer* donde se almacenan las muestras capturadas por el convertidor A/D antes de transferirlas, a través de la interfaz de bus, a la memoria del computador. El objetivo de esta memoria es permitir una mayor velocidad de adquisición del sistema. Junto a estos elementos suelen incluirse otros dispositivos en la tarjeta para incrementar su funcionalidad (tarjetas multifunción). Así, es común que se incluya en la misma tarjeta: convertidores D/A, entradas y salidas digitales y circuitos contadores y de temporización. Las salidas analógicas permiten el control de dispositivos de proceso de control analógico. Los puertos digitales de E/S suelen implementarse con ayuda de algún circuito que actúe de puerto paralelo programable y lo mismo ocurre en el bloque contador/temporizador. La comunicación entre los bloques que constituyen la tarjeta se realiza a través de un bus interno y la conexión con el bus de E/S del computador se realiza a través de la interfaz de bus.

Características y criterios de selección de las TADs

Las especificaciones más importantes de una TAD son:

- **Número de entradas analógicas:** hace referencia al número de canales de entrada disponibles en la tarjeta (4, 8, 16, 32, 64 o más). Se trata de uno de los criterios básicos de selección ya que debe acomodarse a las demandas de nuestra aplicación. En ocasiones, pueden incrementarse los canales añadiendo más de una tarjeta de adquisición al computador, pero es necesario verificar que las tarjetas *dispongan de algún mecanismo para sincronizarse. En el entorno del PC se recomienda no introducir* más de tres tarjetas de adquisición (de tipo PCI) por problemas térmicos, aunque esta limitación se resuelve utilizando un PC industrial que presenta un sistema de refrigeración y de alimentación mejor adaptado. *También puede incrementarse el número de canales recurriendo a multiplexores externos.*

  En la captura de las señales de entrada las tarjetas pueden utilizar un único convertidor A/D compartido entre varias entradas o utilizar convertidores independientes para cada entrada analógica. En el primer caso, se utiliza un multiplexor para seleccionar la entrada correspondiente (figura 25.5). En los casos en que se requiera el muestreo simultáneo puede utilizarse un esquema como el de la figura 25.6 donde se ha conectado un circuito de "S&H" en cada entrada. En cualquier caso, en las tarjetas con convertidor único, la conversión de las señales de entrada se efectúa secuencialmente lo que repercute negativamente en la velocidad máxima de adquisición.

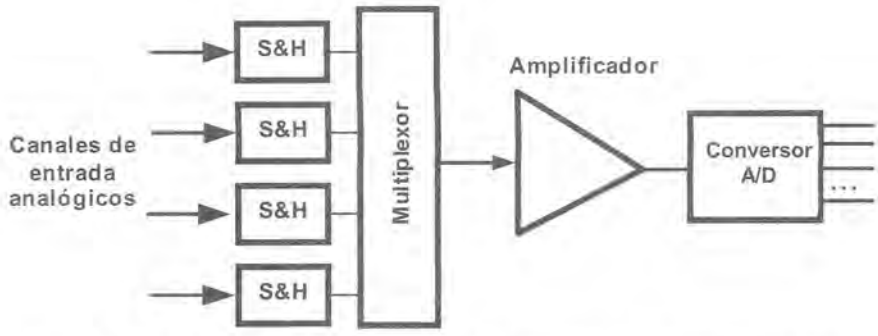


Figura 25.6. Tarjeta de adquisición con un solo A/D y muestreo simultáneo.

Las tarjetas de adquisición que utilizan un conversor independiente para cada canal de entrada presentan mayor coste ya que se multiplica el hardware, pero permiten la captura y conversión simultánea de las señales de entrada, ofreciendo mayores prestaciones (figura 25.7).

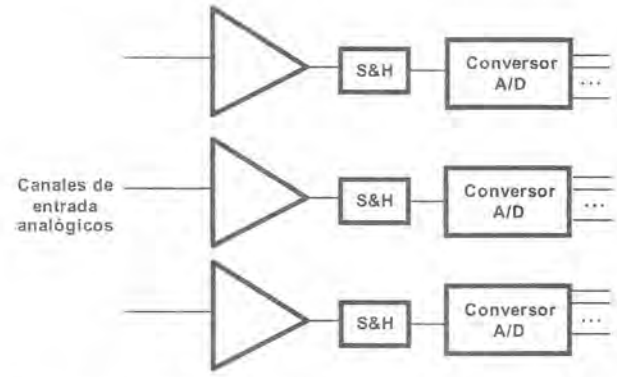


Figura 25.7. Tarjeta de adquisición con convertidores independientes por cada entrada.

- **Configuración de las entradas analógicas.** Las entradas de las tarjetas de adquisición suelen soportar dos configuraciones básicas:
  - **Entradas referidas a masa** (*single-ended inputs*): un terminal de entrada en todos los canales está referido a la masa del sistema de adquisición y, el otro, está disponible para su conexión a entradas exteriores. Esta configuración es aceptable cuando la diferencia de tensión entre la entrada conectada a la masa del sistema de adquisición y el común analógico no sea significativa; de lo contrario se producen errores debido a las corrientes de los bucles de masa. Generalmente se utiliza en la adquisición de señales de nivel alto donde el error introducido por la señal en modo común es despreciable (figura 25.8a).
  - **Entradas en modo diferencial** (*differential input*): en este caso no existe ningún terminal referido a masa. Los dos terminales de entrada se corresponden con los dos terminales de entrada de la tarjeta. Esta configuración es útil en la adquisición de señales de bajo nivel o señales inmersas en entornos ruidosos [el ruido presente sobre ambos terminales es tratado como una señal de modo común y es atenuado por la característica CMRR del amplificador (figura 25.8b)].

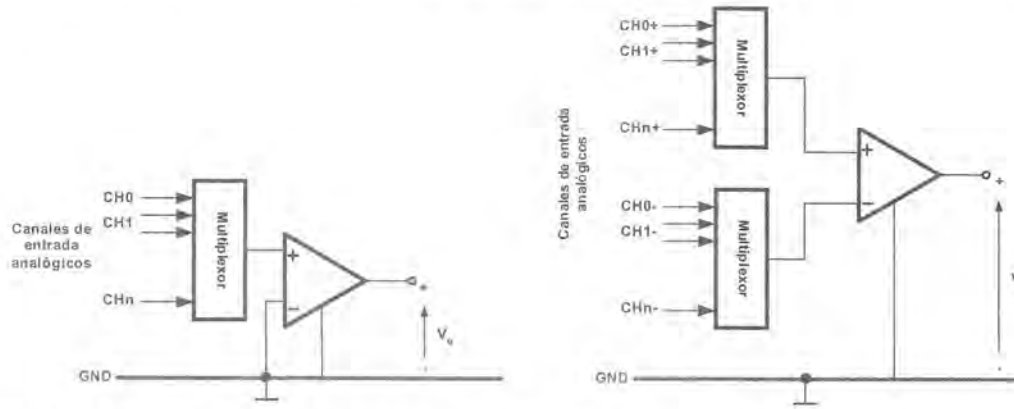


Figura 25.8. Configuraciones de entrada de una tarjeta de adquisición en: (a) referido a masa, y (b) diferencial.

Generalmente, las tarjetas disponen de la posibilidad de configurar sus entradas en estos modos (mediante puentes o por software). El inconveniente del modo diferencial es que el número de canales

de la tarjeta se reduce a la mitad. Relacionado con la entrada también podemos considerar la forma de realizar el acoplamiento de la entrada a la tarjeta. Se especifica como dc o ac, que indican conexión directa o acoplamiento en alterna. El acoplamiento en ac inserta un condensador en serie con la entrada para eliminar la componente de continua.

Las tarjetas de adquisición suelen tener valores de impedancia de entrada muy altos, del orden de 1 a 100 GΩ y los canales de entrada suelen estar dotados de circuitos de protección frente a tensiones excesivas (es importante verificar estos valores para evitar daños en la tarjeta).

Normalmente, las señales de entrada al convertidor requerirán algún tipo de acondicionamiento. Cuando el número de señales a capturar sea elevado se dispone de sistemas especialmente diseñados para resolver esta necesidad que permiten conectar hasta 3000 señales a una sola tarjeta de adquisición. Estos sistemas constan de un conjunto de módulos que se montan sobre un chasis externo y se conecta a una tarjeta controladora que reside en el PC[3].

- **Mecanismo de inicio de captura (disparo):** un aspecto interesante en las tarjetas se refiere a la posibilidad de realizar la adquisición sincronizada con algún evento externo (*triggering* o disparo). Generalmente, la adquisición se inicia por software, escribiendo sobre un bit de un registro determinado de la tarjeta o bien se realiza mediante un temporizador interno a un ritmo establecido. Sin embargo, también puede iniciarse externamente cuando la tarjeta dispone de una entrada especial denominada entrada de **disparo externo** (*trigger input*). Así, cuando una señal toma un valor determinado, por ejemplo, cuando se introduce un flanco de subida, se realiza la captura. El disparo exterior suele utilizarse en aplicaciones donde la temporización entre muestras sea crítica o cuando sólo se pretende capturar datos en momentos exactos de tiempo. Algunas tarjetas permiten mecanismos de disparo más sofisticados como el **disparo analógico** (*analog triggering*) basado en la presencia de determinados niveles analógicos introducidos por la entrada de disparo. Por ejemplo, puede realizarse la captura cuando una señal sobrepasa un valor, se mueve dentro de un margen o incluso pueden adoptarse modos de captura definiendo ciclos de histéresis. La figura 25.9 muestra una posibilidad de disparo basado en el nivel analógico de entrada.

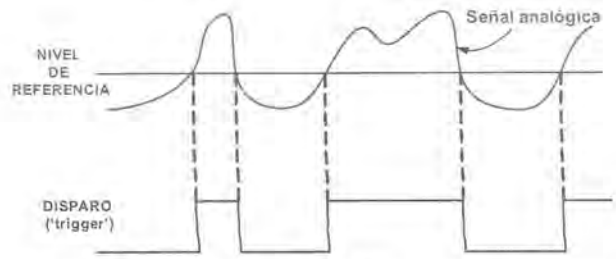


Figura 25.9. Disparo basado en niveles analógicos. Cuando la señal de entrada sube por encima de un valor de referencia, línea a trazos, se produce el disparo representado por un pulso digital en la parte inferior de la figura.

Algunas tarjetas permiten también que el usuario programe modos de rastreo. Un rastreo es una serie de muestreos sobre los canales de entrada repetidos en un orden determinado. Un modo de rastreo muy común es la captura de un número determinado de muestras después del disparo.

[3] Por ejemplo, el bus SCXI de National Instruments.

- **Velocidad de adquisición:** es uno de los parámetros más importantes en la selección de una tarjeta. Especifica la velocidad máxima con que el hardware de adquisición realiza la toma de muestras (frecuencia de muestreo máxima); su inverso es el tiempo de adquisición. Suele especificarse considerando un solo canal conectado, así que debe dividirse por el número de canales para calcular su velocidad real en el caso de realizar un barrido secuencial por todos los canales en una tarjeta de adquisición con un solo convertidor A/D.

  Es importante no confundir este parámetro que especifica el rendimiento global de la TAD, con el tiempo de conversión del convertidor A/D, algunos fabricantes lo usan en su lugar ya que es mucho menor, pues no incluye los retardos asociados al tiempo de establecimiento/conmutación del multiplexor, tiempo de establecimiento del amplificador, del circuito de S&H y los retardos de la transferencia de datos de salida (véase capítulo 23). Tenga en cuenta que a mayor velocidad la TAD tiene mayor coste. Para evitar el submuestreo de la señal, "*aliasing*", algunas tarjetas incorporan filtros *antialiasing* que pueden configurarse en una determinada frecuencia de corte por software o hardware.

  Una frecuencia de muestreo elevada proporciona señales con mayor calidad de definición en el tiempo aunque también aumenta el flujo de datos hacia el procesador. El ancho de banda que proporciona un bus como PCI puede soportar velocidades de muestreo bastante elevadas (1 MS/s), sin embargo, cuando los datos se adquieren a mucha velocidad pueden surgir dificultades para almacenarlos en memoria a la misma velocidad y en el mismo momento que se adquieren. Para solventar este problema las tarjetas de adquisición suelen incorporar memoria intermedia (FIFO *buffers*) donde se almacenan los datos antes de ser transferidos al bus. Esta facilidad permite una mayor velocidad en la adquisición aunque debe gestionarse adecuadamente este *buffer*. En general, se dispone de una capacidad limitada (512 ó 1024 muestras), aunque, en algunas tarjetas, la memoria disponible puede llegar a ser mucho mayor (1 GB).

- **Resolución:** la resolución de convertidor es otro de los parámetros más importantes y debe de elegirse en función de la precisión que se pretenda en la medida. Las TADs suelen utilizar convertidores con resoluciones de 8, 12, 14 ó 16 bits; sin embargo, este dato debe tomarse con precaución (véase capítulo 24).

- **Exactitud:** los fabricantes suelen especificar la exactitud en porcentaje del valor de fondo de escala o en número de bits (0,01% suele corresponder a tarjetas de alta calidad). Este parámetro engloba una combinación de varios errores que están presentes, sobre todo, en el circuito analógico de entrada. Para tener en cuenta el efecto conjunto de todos estos errores y poder conocer la exactitud del sistema de adquisición de una forma global se puede utilizar el ENOB (*Effective Number Of Bits*) que mide la actuación del sistema de adquisición en su conjunto en condiciones reales de funcionamiento, realizando las pruebas a la velocidad máxima de adquisición, usando más de un canal y midiendo una señal dinámica (véase capítulo 23). Podemos encontrarnos con tarjetas de 16 bits con un número efectivo de bits de 13 o menos.

  Algunas tarjetas también disponen de la capacidad para realizar una auto-calibración del convertidor A/D que permite eliminar los errores de *offset* y de ganancia de cada canal de entrada.

- **Margen de entrada:** especifica los límites de tensión de entrada de la tarjeta. Por lo general, se diferencia entre señales unipolares y bipolares. Las señales unipolares admiten únicamente niveles de tensión positivos, por ejemplo 0 a 5 V, mientras que las bipolares permiten las dos polaridades, como –5 a +5 V o –10 a +10 V. Las tarjetas suelen usar puentes para configurar este margen o bien se hace por software. Los fabricantes suelen especificar también el denominado margen dinámico de la tarjeta que es una medida que relaciona la señal más pequeña que puede medirse respecto a la mayor señal de entrada.

Con el objeto de aprovechar todo el margen dinámico del conversor la mayoría de las tarjetas de adquisición incorporan un circuito amplificador de ganancia programable. El usuario puede programar la ganancia necesaria en función de la amplitud de la señal de entrada analógica. En algunas tarjetas esta ganancia es común a todos los canales y, en otras, puede establecerse una ganancia independiente para cada canal.

- **Tipo de convertidor A/D:** el tipo de convertidor utilizado depende de las prestaciones de la tarjeta. Las tarjetas de prestaciones medias con bajas resoluciones suelen incorporar un convertidor de aproximaciones sucesivas. En tarjetas de prestaciones medias con altas resoluciones (> 16 bits) suelen utilizar conversores delta-sigma.
- **Transferencias de datos:** la forma en que se realiza la transferencia de datos entre la tarjeta y la memoria del computador es una característica que puede llegar a ser crítica. En general, las TADs disponen de varias formas de realizar esta transferencia: programada, interrupciones, DMA.

  Cuando se inicia un proceso de conversión, los datos capturados se almacenan de forma temporal en la memoria interna de la tarjeta hasta que se lean y pasen a la memoria interna del computador. Como el *buffer* tiene un límite de capacidad, el programa de usuario debe leerlo periódicamente antes de que se llene o, de lo contrario, se perderían datos. La tarjeta dispone de un bit de estado que informa cuando existen datos en el *buffer* dispuestos para ser leídos y cuando está vacío. El programa del usuario puede monitorizar constantemente este bit para saber cuándo debe leer los datos y situarlos en la memoria del computador; este tipo de transferencia se denomina **E/S programada** y es un proceso costoso en tiempo de CPU, ya que la CPU está siempre ocupada comprobando el bit de estado del *buffer*.

  Otra alternativa es dejar que sea la tarjeta de adquisición la que advierta al computador cuando tenga datos disponibles y sea entonces cuando se proceda a leer el *buffer*. Este proceso se realiza mediante la generación de una **interrupción** a través de una línea disponible en el sistema de interrupciones del computador. Mediante este mecanismo, se libera a la CPU de una dedicación constante de monitorización, aunque debe tenerse presente el tiempo que representa la atención de la interrupción (latencia de interrupción). Este retraso puede resultar excesivo en determinadas aplicaciones en tiempo real.

  Cuando las transferencias a la memoria del computador deban realizarse a alta velocidad, se opta por usar el DMA. En este caso, un procesador especializado, no la CPU, toma el control de los buses del PC y realiza directamente el direccionamiento de la memoria del computador pudiendo transferirse bloques enteros de datos. La CPU queda liberada para realizar otras tareas. En el caso de tarjetas de adquisición que funcionen con altas frecuencias de muestreo es importante que posean este modo de transferencia de datos.

Existen otros factores importantes en las tarjetas de adquisición relacionados con parámetros dinámicos. Así, es necesario considerar las derivas térmicas y el ruido del sistema, sobre todo, en las tarjetas de alta resolución. Para minimizar estos errores, las tarjetas utilizan componentes de calidad con valores muy estables en amplios márgenes de temperatura y/o hacen uso de técnicas de compensación. En algunas se incorpora un sensor para conocer la temperatura real de la tarjeta asegurando medidas más fiables si se conoce su tolerancia con la temperatura.

Las tarjetas suelen requerir un tiempo antes de realizar medidas para que la temperatura se estabilice, aproximadamente 15 minutos. Para reducir el ruido del sistema las tarjetas utilizan componentes de alta calidad y un diseño especial con partes apantalladas y varios planos de masa. En el caso de tarjetas de adquisición montadas en los *slots* de un PC es aconsejable alejarlas en lo posible de la tarjeta de vídeo y dejar, si es posible, un *slot* libre a cada lado para reducir el ruido.

Tarjetas de adquisición multifunción

Las tarjetas de adquisición suelen incluir otras prestaciones para incrementar su funcionalidad: salidas analógicas, E/S digitales y circuitos contadores y de temporización.

- **Salidas analógicas:** existen tarjetas de adquisición que proporcionan al usuario varios canales de salida analógicos. Las salidas analógicas pueden utilizarse para el control de ciertos dispositivos de una planta. A la hora de seleccionar es interesante considerar: la resolución del convertidor D/A, el número de canales de salida y la gama de tensiones de salida. Generalmente se proporcionan varios canales analógicos que se obtienen desde convertidores D/A independientes. La corriente que puede suministrar el canal analógico es limitada (5÷10 mA) por lo que debe tenerse en cuenta si se pretende excitar cargas considerables. En cuanto a la gama de salida, suele utilizarse un convertidor D/A multiplicador de forma que la salida depende de la referencia que se utilice, que puede ser interna o externa. La gama típica de salida es de ± 10 V.
- **Entrada/salida digital:** se utilizan para control de procesos, activación de alarmas, comunicación con periféricos, recepción de señales de sensores discretos, generación de señales compatibles TTL, etc. Para elegir una tarjeta en que se precise esta característica es interesante considerar: el número de canales de entrada/salida, su configuración y compatibilidad eléctrica. El número de canales especifica el número de bits disponibles para entrada y salida (a menudo los bits de E/S pueden configurarse individualmente o configurarse como puertos digitales de forma que pueden ser leídos o escritos simultáneamente). También puede considerarse la velocidad con la que pueden transmitirse los datos y la capacidad de control de otros dispositivos con estos puertos. En algunas tarjetas es posible establecer ciertos protocolos de comunicación a través de estas líneas para la comunicación y/o control de un periférico externo. Los niveles de entrada y salida digital suelen ser compatibles con niveles TTL y/o CMOS. Es importante, no obstante, comprobar las capacidades máximas de corriente para verificar la carga máxima que pueden manejar directamente y los valores máximos admitidos para no dañar la tarjeta.
- **Contadores/Temporizadores:** las funciones del contador/temporizador se emplean normalmente para activar conversiones, generar interrupciones, contar sucesos, generar pulsos temporizados, generar ondas cuadradas, medidas de frecuencia, de anchura de pulsos, etc. Las tarjetas suelen incorporar algún circuito programable universal que realiza las funciones de temporizador/contador. Es interesante comprobar el número de temporizadores independientes, su frecuencia máxima de trabajo (el *timer* —temporizador/contador— se gobierna por un reloj interno en la tarjeta) y el número de bits que utilizan los registros del contador para comprobar que se adapte a nuestra aplicación.

25.3 Buses de instrumentos

Los sistemas de adquisición de datos basados en instrumentos independientes pueden funcionar de forma autónoma o bien conectarse a un computador y a otros instrumentos a través de un bus de comunicación. En general, el bus más difundido para la conexión de instrumentos es el bus GPIB. En el caso de los instrumentos modulares, la interconexión entre los módulos del sistema se realiza a través de un bus de conexión local de altas prestaciones como VXI o PXI. Este bus adopta la forma de un panel posterior al que se conectan los distintos módulos que componen el sistema de adquisición montado en un *rack* (figura 25.4). Estos sistemas pueden conectarse, a su vez, a computadores, a otros *racks* de módulos o a instrumentos independientes permitiendo la configuración de soluciones muy versátiles. A continuación, veremos las características básicas de estos buses.

25.3.1 EL BUS GPIB

Introducción

El GPIB (*General Purpose Interface Bus*) fue desarrollado en los años sesenta por *Hewlett-Packard* (se le llamaba por entonces HP-IB) para la conexión y el control de los instrumentos programables de este fabricante. Con la proliferación de instrumentos programables y de computadores digitales surgió la necesidad de disponer de un bus estándar de alta velocidad que permitiese la comunicación entre equipos de diferentes fabricantes. En 1975 el IEEE publicó el estándar ANSI/IEEE 488-1975 (*IEEE Standard Digital Interface for Programmable Instrumentation*) que contenia las especificaciones eléctricas, mecánicas y funcionales de la interfaz para la interconexión de instrumentos. Este documento se revisó posteriormente en 1978 (IEEE 488.1) y terminó siendo ampliamente adoptado en la industria bajo tres denominaciones diferentes:

- GPIB (*General Purpose Interface Bus*)
- *Hewlett-Packard Interface Bus* (HP-IB)
- IEEE 488 Bus
- IEEE 488.1 (esta denominación se adoptó con posterioridad)

Con posterioridad, en 1987, se publicó un suplemento al estándar IEEE 488 de 1978 denominado: "*IEEE 488.2 Codes, Formats, Protocols and Common Commands*" que añadía consideraciones sobre la sintaxis y el formato de los comandos del estándar con vistas a incrementar la compatibilidad entre los sistemas. Este nuevo estándar no reemplazó al IEEE-488.1 original, de hecho, muchos dispositivos siguen todavía usándolo. En 1990, la especificación IEEE-488.2 incluyó el documento: "*Standard Commands for Programmable Instrumentation (SCPI)*". SCPI define un conjunto específico de comandos para el control y la configuración de cada clase de instrumento. Mediante SCPI se garantiza una compatibilidad total entre instrumentos de distintos fabricantes siempre que sigan el estándar.

A continuación, definiremos las características físicas de este bus (descripción del conjunto de cables y conectores que transportan las señales eléctricas) y haremos algunos comentarios sobre su lógica de funcionamiento, no obstante, se debe recurrir a la bibliografía proporcionada para una descripción detallada de este estándar.

Especificaciones físicas

Las especificaciones físicas se refieren a la configuración del bus, longitud, tipo de conector y cable utilizados. GPIB es un bus digital paralelo de 8 bits, con velocidades de transferencia de 1 MB/s[4] y consiste físicamente en un cable apantallado (figura 25.10) con dos conectores que permite conectar múltiples dispositi-

Figura 25.10. Cable GPIB.

[4] Existen versiones de alta velocidad a 8 MB/s (HS488).

vos entre sí en forma lineal o en estrella (figura 25.11). Los conectores son de tipo americano o europeo. El americano tiene 24 pines y el europeo 25 (IEC625.1) y el cable utiliza dos conectores en cada extremo de tipo macho/hembra, apilable (*piggyback*) (figura 25.10). Los instrumentos utilizan una base de tipo hembra.

La longitud máxima del bus es de 20 m y el número máximo de dispositivos conectados es de 15, siendo uno de ellos un dispositivo especial llamado **controlador** (suele ser una tarjeta conectada al bus interno del PC). Al menos las dos terceras partes de los dispositivos conectados deben estar encendidos. En el caso del estándar IEEE 488.1 la longitud del bus se reduce a 15 m y todos los dispositivos deben estar encendidos. La distancia máxima entre dos dispositivos es de 4 m. Las limitaciones de distancias pueden evitarse añadiendo repetidores. Los niveles de tensión que se utilizan son TTL con lógica negativa. El "1" lógico, se asocia con un nivel TTL bajo (<0,8 V) y el "0" lógico con un nivel TTL alto (>2,0 V).

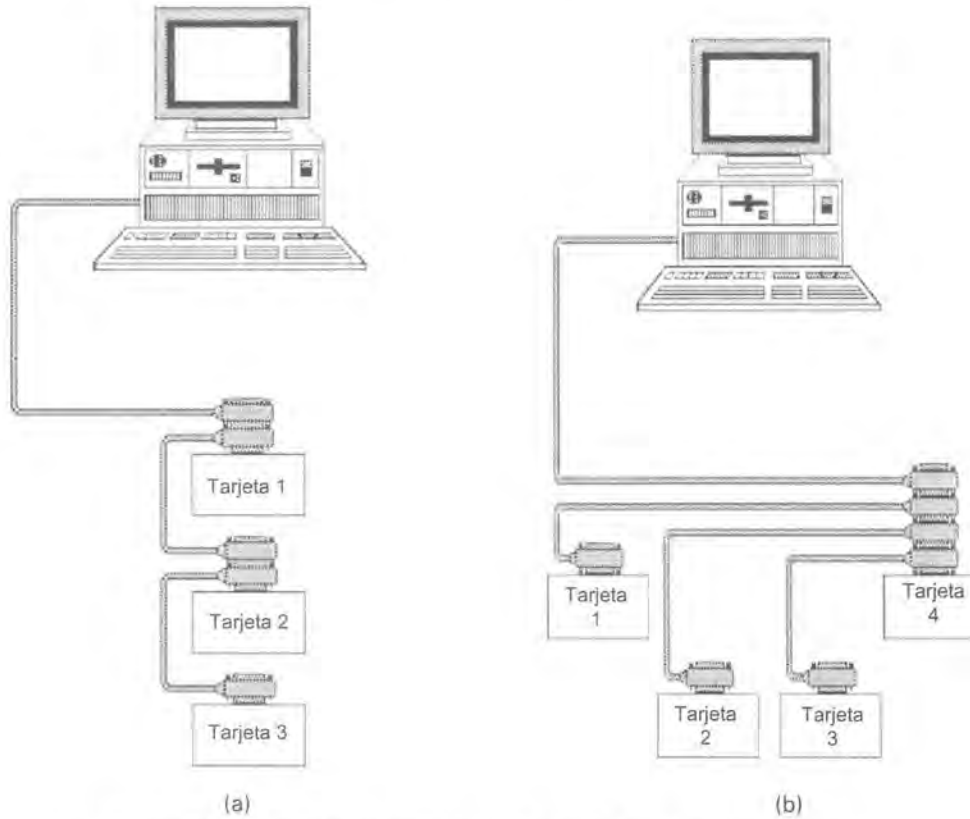


Figura 25.11. Topología del bus GPIB: (a) en bus y, (b) en estrella.

Especificaciones funcionales

La comunicación a través del bus IEEE-488 se realiza considerando tres tipos de elementos: **controladores, *talkers*** (emisores) y ***listeners*** (receptores).

Un *talker* es un dispositivo que emite información sobre el bus y un *listener* el dispositivo que la recibe. La misión del controlador es dirigir el flujo de datos asignando por anticipado qué dispositivo del bus funciona como *talker* (sólo puede ser uno) y cuál o cuáles como *listener* (pueden haber varios en un momento dado). El controlador suele ser una tarjeta que se instala en el PC.

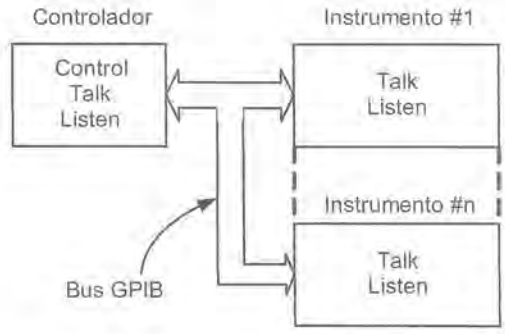


Figura 25.12. Concepto de bus IEEE- 488.

Puede haber varios controladores pero sólo uno está activo en cada momento (se le denomina CIC, *Controller in Charge*). La mayoría de los instrumentos pueden comportarse como *talker/listener* aunque algunos dispositivos son siempre sólo *talker* o sólo *listener*. Por ejemplo, una impresora es sólo *listener*.

La comunicación en el GPIB se realiza por medio de mensajes que se envían a los dispositivos. Un dispositivo se compone de dos partes (figura 25.13) y puede recibir dos tipos de mensajes:

- **Mensajes específicos del dispositivo**, que se dirigen a las funciones del instrumento. Los mensajes específicos del dispositivo se denominan datos o **mensajes de datos** y contienen información específica del dispositivo como: instrucciones de programación, resultados de medidas o estado del instrumento.
- **Mensajes de control del medio**, que se dirigen a las funciones de la interfaz. Los mensajes de control de medio son los encargados de gestionar el sistema de comunicaciones y normalmente se denominan comandos o ***mensajes comando***. Estos comandos se utilizan para inicializar el bus, seleccionar/direccionar dispositivos o establecer los modos de funcionamiento para la programación del instrumento (local o remota).

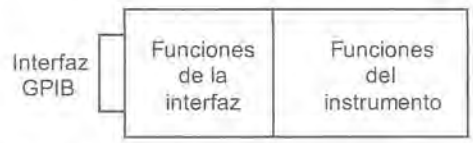


Figura 25.13. Instrumento IEEE-488.

El controlador del GPIB (generalmente la tarjeta dentro del PC) y cada instrumento sobre el bus debe tener una única identificación, su dirección, especificada físicamente por medio de un conjunto de interruptores en el propio dispositivo. Los dispositivos tienen una dirección primaria que es un número comprendido entre 0 y 30 (aunque sólo puede haber 14 dispositivos en el bus). La 0 o la 21 suelen asignarse al controlador y el resto puede asignarse a los dispositivos. El controlador usa esta dirección para construir un mensaje de direccionamiento que indica el dispositivo al que va dirigida si se debe ponerse en estado de lectura (esperar la llegada de datos) o de escritura. Aunque puede haber varios instrumentos conectados al bus, sólo uno reconoce la secuencia de direccionamiento comparando la dirección que tiene asignada físicamente (mediante los interruptores) con la dirección que circula por el bus. El bus consiste en 16 líneas de señales y 8 líneas de tierra. Las 16 líneas de señales se dividen en 3 grupos:

– 8 líneas de datos.
– 3 líneas de para el control de la transferencia (*handshake*).
– 5 líneas de control del bus.

Figura 25.14. Señales de bus GPIB.

Las líneas de datos (DIO7...DIO0) son bidireccionales y transportan mensajes en código ASCII de 7 bits. La información transferida puede contener mensajes de direccionamiento, órdenes de programación, medidas de un instrumento o información sobre el dispositivo.

Las líneas de control de transferencia coordinan las transferencias de datos por el bus y aseguran que no se emita si el receptor no está preparado para recibir o que la transmisión dure lo suficiente para que se acomode al dispositivo más lento. Constan de las siguientes señales:

- **DAV** (*Data Valid*): la controla el emisor e indica que sobre el bus hay un dato correcto.
- **NRFD** (*Not Ready For Data*): esta línea la controlan los receptores para informar al emisor si están o no preparados.
- **NDAC** (*Nor Data ACepted*): la utilizan los receptores para señalar si han aceptado o no el dato presente en el bus.

Las líneas de control del bus gestionan su funcionamiento y las señales básicas son:

- **IFC** (*InterFace Clear*): la gobierna el controlador de bus y permite inicializar el bus a un estado conocido.
- **ATN** (*ATteNtion*): según su nivel lógico los valores que circulan por el bus se consideran como mensajes de interfaz o mensajes de instrumento.
- **REN** (*RemoteENable*): la gobierna el controlador y permite, con otros mensajes de interfaz, seleccionar el modo de control local o remoto de un dispositivo.
- **EOI** (*End Or Identity*): indica el fin de transferencia de datos si la usa un *talker* o identificación si la usa un controlador.
- **SRQ** (*Service ReQuest*): la activa un instrumento para notificar al controlador del bus que necesita sus servicios. Cuando la detecta el controlador, efectúa una encuesta para identificar al instrumento que ha solicitado la petición.

La tabla 25.1 resume las especificaciones del bus GPIB.

Especificación	Concepto	Descripción
Mecánicas	topología	Estrella o bus
	longitud	2 m entre instrumentos, máximo 20 m
	conector	Instrumento (Hembra) Cable (apilable: macho-hembra)
Eléctricas	"1" lógico	< 0,8 V
	"0" lógico	> 2 V
Funcionales	Tipos de instrumentos	*Controller* (controlador) *Talker* (emisor) *Listener* (receptor)
	Nº máx de instrumentos	15
	Nº instrumentos activos	> 2/3
	Velocidad	1 MBps
	Dirección en instrumento	Conmutadores o memoria pasiva

Tabla 25.1. Especificaciones básicas de la norma IEEE 488.

25.3.2 BUS VXI

Físicamente, VXI (*VMEbus eXtension for Instrumentation*)[5] consiste en un chasis con un plano posterior (*backplane*) sobre el que se conectan unos módulos en forma de tarjetas enchufables (figura 25.4); con este sistema puede configurarse una solución modular de instrumentación.

VXI se utiliza fundamentalmente cuando se necesita un sistema de adquisición de datos fiable, de altas prestaciones (alta velocidad de adquisición), con gran número de variables a capturar y con posibilidades de ampliación. En general, para la adquisición de pocos canales (hasta 20) una tarjeta de adquisición de datos puede ser suficiente. Para un número de canales superior (hasta 100) puede utilizarse un instrumento externo de adquisición independiente. Cuando las necesidades aumentan, VXI puede ser la mejor solución. Por ejemplo, con un bastidor de 6 ranuras pueden disponerse de hasta 320 canales. En general, el coste de las tarjetas es menor que el de un instrumento independiente y su potencia es superior a la de una tarjeta de adquisición. Las ventajas del sistema pueden resumirse en [25.5]:

a) **Sistema abierto** (IEEE-1014), **flexible y modular**. La arquitectura está soportada por numerosos fabricantes de instrumentos siendo posible elegir entre cientos de productos para satisfacer las más variadas necesidades (convertidores A/D, acondicionadores de señal, salidas analógicas y digitales, generadores de formas de onda, etc.). Es posible añadir más módulos a un *rack* para dotar al sistema de mayores prestaciones, conforme lo demande nuestro sistema de adquisición de datos. Incluso, ampliar a un segundo bastidor de tarjetas utilizando un cable estándar y un protocolo denominado MXIbus. En general, VXI es una solución muy adecuada para cualquier sistema con un número de canales comprendidos entre 10 y 1000.

b) **Alto rendimiento** al estar basado en un bus de 32 bits de altas prestaciones. Puede transferir datos a una velocidad teórica de 10 MB/s (GPIB tiene un límite de 1 MB/s). En caso de que se desee comunicación entre tarjetas, VXI dispone de un bus local que puede llegar a 100 MB/s. Se puede compartir datos entre tarjetas o enviar información a una unidad de disco incorporada.

[5] Derivado de VMEbus, *Versa Modular European*, un estándar de planos posteriores (*backplane*) de computadores.

c) **Fiable y robusto**. Se aseguran medidas precisas y fiables gracias a las protecciones y al apantallamiento que establece límites estrictos sobre las interferencias por conducción y radiación. Las tarjetas VXI son más fiables que las tarjetas de adquisición de datos destinadas a un PC. Suelen estar rodeadas de una pantalla metálica para evitar interferencias y permiten mayores resoluciones y características superiores a las tarjetas de adquisición.

d) **Tamaño reducido** que favorece la movilidad, portabilidad y su proximidad al proceso a controlar.

e) **Altas velocidades de adquisición** y mecanismos precisos de temporización y disparo entre todas las tarjetas del *rack*.

f) **Configuración y programación sencilla** mediante varios lenguajes de programación.

Un sistema VXI puede configurarse de tres formas:

a) Un módulo especial (GPIB-VXI) hace de puente entre el bus VXI y una tarjeta GPIB que actúa de controlador dentro del PC.

b) Un módulo funciona como un computador dentro del rack, al que puede conectarse teclados y pantallas. Esta configuración permite reducir el tamaño del sistema y proporciona un alto rendimiento al estar conectado directamente al bus VXI.

c) Un módulo interfaz (*MXI, Multi-system eXtensión Interface*) dentro del rack permite utilizar un ordenador externo como controlador del bus VXI. Esta configuración es funcionalmente equivalente al método anterior, aunque más flexible.

Otra alternativa de instrumentación modular muy difundida es **PXI**[6]. PXI utiliza una variante del popular bus PCI, muy utilizado en los computadores personales, para la interconexión de los módulos. A diferencia del bus PCI, esta arquitectura dispone de unas características de sincronización avanzadas, pudiendo transmitirse por el bus señales digitales (TTL) a alta velocidad e incluso señales analógicas entre los distintos módulos. Son posibles dos configuraciones: un chasis con módulos funcionando como esclavos y conectados a un computador externo o un chasis con un módulo que actúa como computador del sistema.

Resumen

En este capítulo se estudian distintas alternativas para configurar un sistema de adquisición de datos. Básicamente, se concreta en tres tipos o en una combinación de éstos:

1º Sistemas basados en un computador (PC) junto a una tarjeta de adquisición de datos.
2º Instrumentos de adquisición independientes.
3º Instrumentos de adquisición modulares.

Los sistemas de adquisición basados en un computador se han convertido en una de las alternativas más prometedoras debido a su coste y versatilidad. En esta configuración, suele utilizarse un PC como plataforma para realizar el procesamiento, el almacenamiento y la visualización de datos y una tarjeta de adquisición, conectada directamente al bus del computador, para la captura de las señales analógicas.

Los sistemas de adquisición basados en instrumentos independientes suelen utilizarse en aplicaciones más exigentes que requieran la captura de un gran número de variables con gran precisión y fiabilidad. En general, son soluciones de mayor coste, pero están mejor acondicionadas y ofrecen mayores prestaciones que la anterior. A menudo, estos instrumentos se conectan entre sí a través de un bus como el GPIB.

[6] *PCI eXtension for Instrumentation*, extensión de PCI para instrumentación de *National Instruments*.

Los sistemas basados en instrumentos modulares permiten configurar sistemas de adquisición más potentes y a un coste que puede llegar a ser más reducido que con instrumentos autónomos, sobre todo, en sistemas de cierta dimensión o con necesidad de crecimiento futuras. Los sistemas basados en módulos son muy robustos, muy flexibles y con gran potencial de procesamiento. Estos instrumentos se configuran a través de módulos conectados en un *rack* compartiendo un bus común. Entre los buses más utilizados en este tipo de instrumentos modulares está el VXI.

26

Buses de campo

26.1 Introducción: conceptos previos
26.2 Características de las redes de control
26.3 Buses de campo

26.1 Introducción: conceptos previos

La comunicación se concibe como un intercambio o flujo de información entre los elementos que componen un sistema. La comunicación es una tarea esencial en la industria porque cada vez se utilizan más automatismos de control y dispositivos de medida que deben coordinarse para realizar un trabajo útil. En general, la comunicación moderna entre dispositivos se realiza por medio de las denominadas **redes digitales**. Una red está constituida básicamente por tres elementos: un conjunto de terminales o nodos que intercambian información, un medio físico de enlace que conecta estos nodos y un software que permite su gestión.

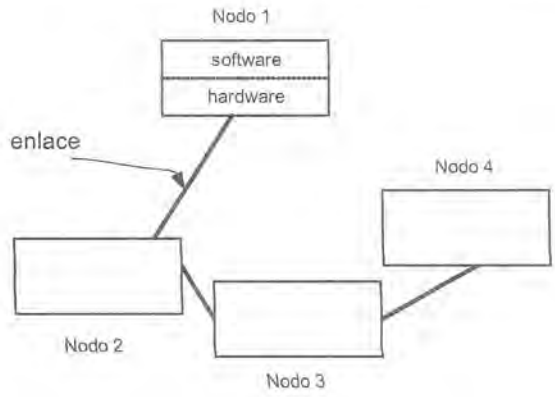


Figura 26.1. Elementos de una red de comunicaciones.

La información se transmite en general de dos formas:

- **Serie:** la información se transmite bit a bit por una línea hasta completar una palabra (figura 26.2a).

- **Paralelo:** se transmiten palabras completas, es decir, los datos se transmiten sobre las líneas correspondientes de forma simultánea. Existe una línea por cada bit de datos, además de otras líneas dedicadas al control (figura 26.2b).

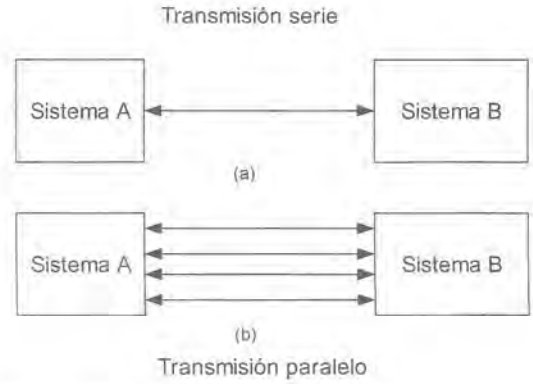


Figura 26.2. Transmisión: (a) serie y (b) paralelo.

La información se transmite en serie en la mayor parte de las redes de comunicación por razones de coste de la conexión sobre todo cuando la distancia a cubrir es elevada. Cuando la velocidad es un condicionante importante y el enlace es de pequeña extensión, puede utilizarse un enlace paralelo. Por ejemplo, en enlaces complejos donde la distancia a cubrir es pequeña (varios centímetros) suele trabajarse en paralelo.

Otros conceptos interesantes relacionados con la comunicación de datos son los siguientes:

- **Conexión punto a punto:** conexión en la que intervienen sólo dos terminales, uno a cada extremo de la línea de conexión.

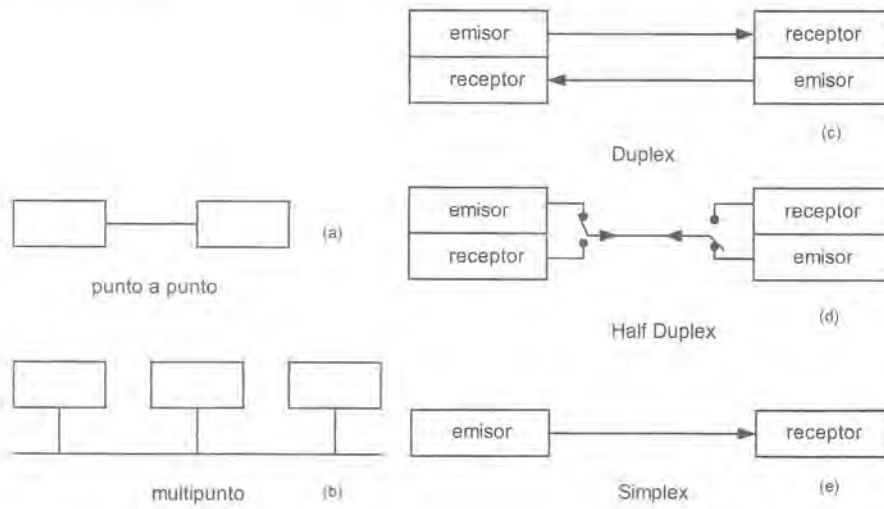


Figura 26.3. Tipos conexión: (a) punto a punto, (b) multipunto, (c) *full-duplex*, (d) *half-duplex*, (e) *simplex*.

- **Conexión multipunto:** conexión de más de dos terminales a través de la misma línea de conexión.

- ***Full-duplex***: comunicación que permite el flujo de datos entre dos terminales en los dos sentidos simultáneamente.
- ***Half-duplex***: comunicación que permite el flujo de datos entre dos terminales en los dos sentidos pero no de forma simultánea.
- ***Simplex***: comunicación que permite el flujo de datos entre dos terminales en un solo sentido.

Redes Industriales

La necesidad de comunicación en la industria viene de la mano de las exigencias que introdujo, fundamentalmente, el **control distribuido**. En sus inicios, el control industrial se realizaba utilizando dispositivos de tipo analógico al que llegaban las señales del proceso mediante el clásico bucle de corriente. Posteriormente, la introducción del computador cambió el panorama: al principio se le encomendaron al computador exclusivamente tareas de supervisión, aunque paulatinamente fue asumiendo las tareas propias del control debido a su flexibilidad y capacidad para la implementación de los algoritmos de control. En este escenario, el control se realizaba de forma centralizada mediante un "gran" computador al que llegaban todas las variables de proceso y del que partían todas las señales de mando (figura 26.4).

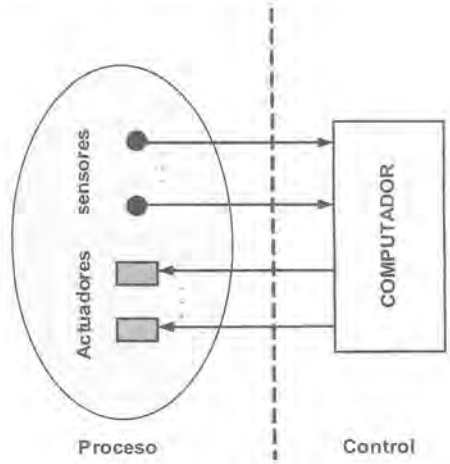


Figura 26.4. Sistema centralizado de control mediante computador

Conforme fueron introduciéndose los dispositivos digitales programables de reducidas dimensiones y bajo coste (microprocesadores, microcontroladores, PLC y DSP) se plantea la posibilidad de controlar, de forma independiente, partes concretas de un proceso formando islas de automatización independientes. Las ventajas de esta opción distribuida se encuentran en la flexibilidad, en la seguridad y en la reducción de coste frente a una solución centralizada. Piense, por ejemplo, que un fallo en el computador central puede dar lugar a una parada total de la planta. No obstante, el control descentralizado origina otra necesidad: la de comunicar todos estos elementos de control ya que deben coordinarse para realizar un trabajo útil. Esta necesidad de comunicación se incrementó aún más con la introducción de los denominados **sensores inteligentes** que permiten la distribución de los datos directamente a través de la red de comunicaciones. Este es el contexto en el que se desarrollaron las redes digitales en la industria. Sin embargo, las comunicaciones en este entorno no se limitan sólo al nivel de planta o de control, también existen sistemas de información encargados de la administración y de la planificación de la producción que resultaría interesante integrar con los sistemas de control. Este concepto de comunicaciones integradas en todos los niveles de la empresa nos lleva a un modelo basado en una estructura jerárquica como el que se describe en la figura 26.5.

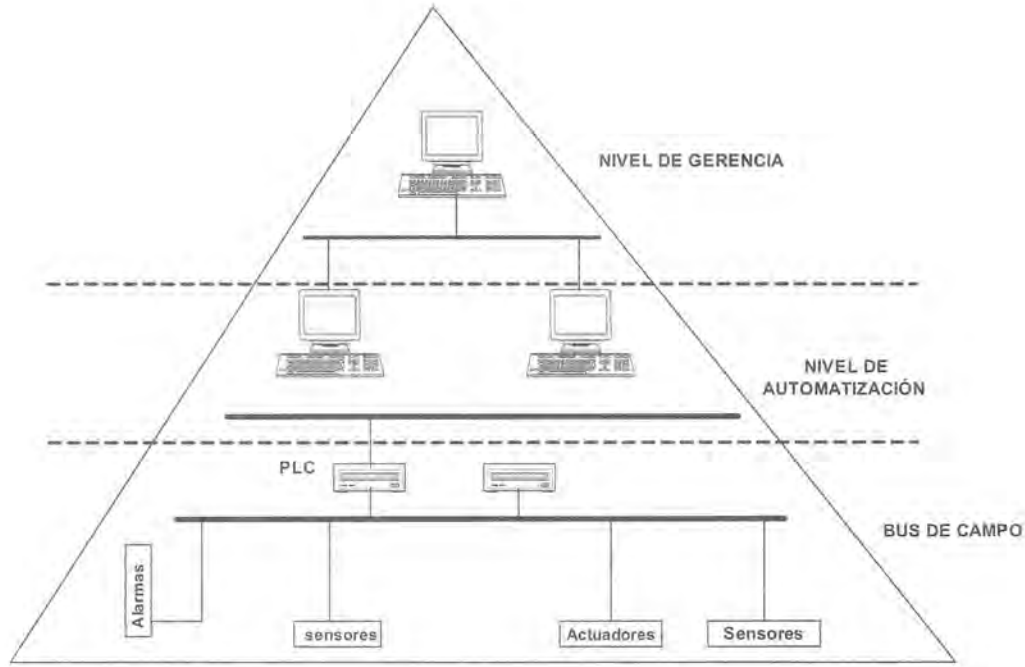


Figura 26.5. Estructura jerárquica de las comunicaciones en la industrial: modelo CIM.

En la parte alta de la pirámide o **nivel de gerencia** se realizan las tareas de gestión, fijando la política de producción en función de recursos y costes. Las redes de comunicación en este nivel tienen desde hace tiempo una fuerte implantación siguiendo estándares muy consolidados. La parte baja es la que está ligada al proceso, a la producción. Los dispositivos en este nivel se comunican por medio de redes digitales denominadas **buses de campo**, siendo en este nivel donde se realiza el control directo de la planta: se adquieren datos de los sensores, se activan alarmas, actuadores y se transmiten los mensajes al nivel intermedio, el **nivel de automatización**, que realiza la supervisión y establece los algoritmos de control. La finalidad última de esta estructura piramidal es poder actuar sobre la producción desde el nivel de gerencia. La implementación de esta estructura define el denominado **modelo CIM** (*Computer Industry Manufacturing*) o fabricación industrial por computador.

Las redes de la parte alta de la pirámide se denominan **redes de datos** y las de la parte baja se denominan **redes de control**. Las **redes de datos** se utilizan fundamentalmente en el ámbito de la gestión y resuelven las necesidades de comunicación entre computadores locales o remotos cuya misión básica es la administración de la empresa. Estas redes están dedicadas especialmente al intercambio y al procesamiento de datos y desarrollaron con fuerza debido al gran volumen de negocio que mueven y al tiempo que llevan funcionando. En general, son redes digitales de tamaño medio (LAN) o de gran tamaño (WAN) y suelen seguir estándares muy conocidos. Las redes de datos están sujetas a un tráfico de información peculiar formado, en general, por grandes volúmenes de datos que se mueven esporádicamente entre los nodos de la red y no suelen ser redes críticas.

Las redes que se utilizan en la parte baja de la pirámide, redes ligadas al proceso o **redes de control** presentan otras características. En el ámbito de este texto estaremos interesados especialmente en este tipo de redes.

26.2 Características de las redes de control

Las **redes de control** están dedicadas al intercambio de la información proveniente de las variables de proceso y presentan características diferentes de las redes de datos que pueden resumirse en [26.2]:

- **Velocidad de respuesta:** en aplicaciones de control la velocidad de respuesta ante un acontecimiento o evento es un factor esencial. A menudo, en las redes de control nos encontramos con sistemas que trabajan en **tiempo real** por lo que se demanda una red que satisfaga esta necesidad, es decir, que dé una respuesta en un tiempo determinado y predecible. Estos requerimientos no son necesarios, en general, en las redes de datos.
- **Prioridades:** estas redes requieren la implementación de un sistema de prioridades de manera que, cuando el flujo de información sea muy intenso, se permita la comunicación a procesos que por su importancia no puedan esperar, por ejemplo, las alarmas en un sistema de control.
- **Tipo de tráfico:** el tráfico de información consiste en las señales de proceso obtenidas por los sensores y señales de control dadas por los sistemas supervisores. En general, este tipo de tráfico está formado por paquetes de datos de pequeño tamaño que se transmiten de forma constante por la red.
- **Seguridad:** esta característica es muy importante en las redes de control, se refiere a su capacidad para la **detección** y **recuperación de fallos**. A menudo las redes de control son redes críticas en el sentido en que un fallo en la red puede originar resultados catastróficos. En general, en una red de datos un fallo puede no tener consecuencias. En procesos críticos es habitual disponer de sistemas redundantes que entran en funcionamiento en caso de avería. También es importante la monitorización de la red y su capacidad de diagnóstico.
- **Inmunidad al ruido:** en general las redes de control discurren por entornos ruidosos por lo que es importante dotarlas de medios de transmisión adecuados.
- **Interoperabilidad:** un objetivo primordial en estas redes es utilizar sistemas de distintos fabricantes. La interoperabilidad entre dispositivos se garantiza si estos siguen un estándar; por desgracia, las redes en este ámbito están sujetas a más incompatibilidades que las redes de gestión y, a menudo, se encuentran sistemas propietarios.
- **Administración:** en las redes de control suele ser frecuente las reconfiguraciones de los nodos y el añadido de nuevos nodos por lo que resulta interesante disponer de esta posibilidad. Por tanto, la **capacidad de mantenimiento** y **administración** de la red es otro punto a considerar.

La tabla 26.1 resume las características básicas de estas redes.

Parámetro	Redes de datos	Redes de control
Ámbito	gestión	proceso
Tamaño	LAN - WAN	LAN
Volumen de datos	grande	pequeño
Procesos críticos	No	Sí
Velocidad	No en tiempo real	Alta (tiempo real)
Crítica	No	Sí
Prioridades	No	Sí
Estandarización	Alta	Baja
Seguridad	Media	Alta
Administración	Sí	Sí / frecuentes reconfiguraciones

Tabla 26.1. Características de las redes de datos y redes de control.

En todas las redes, y en especial en las redes de control, existen un conjunto de factores que deben considerarse a la hora de contrastar su idoneidad para una determinada aplicación. Entre los aspectos más importantes destacan: **la topología de la red, el medio físico** utilizado y el método de **control de acceso al medio**. A continuación, analizaremos estos aspectos.

Topología de redes

La forma en que se interconectan los elementos de una red define la topología de la red. La topología tiene un impacto considerable en su comportamiento. Las topologías básicas en redes locales son tres (figura 26.6): estrella o jerárquica, anillo, bus.

La elección de una u otra topología en las redes de control suele hacerse teniendo en cuenta factores como coste y tolerancia a fallos:

Anillo: cada nodo se enlaza con el siguiente formando un lazo cerrado. Como la información circula por todas las estaciones que forman el anillo su rendimiento está en función de la estación más lenta. La comunicación entre dos nodos se realiza pasando previamente por todos los nodos intermedios que reconocen el mensaje y, si no son los destinatarios, lo retransmiten a la siguiente estación. Como desventaja, si una estación presenta una deficiencia puede verse afectada toda la red.

Estrella: las estaciones se conectan a un nodo común que sirve como punto de enlace entre el resto de nodos. La ampliación de esta red resulta muy sencilla y, si un nodo falla, tan sólo se interrumpe la comunicación con él. El problema se plantea si falla el nodo central por lo que la fiabilidad de este nodo debe ser grande.

Bus: los nodos se unen entre sí a través de un conjunto de líneas comunes. Este tipo de topología es la más habitual en las redes de autómatas y buses de campo por ser muy flexible para incluir nuevos nodos y también por su fiabilidad. No obstante, la fiabilidad depende de cómo se configure desde un punto de vista lógico.

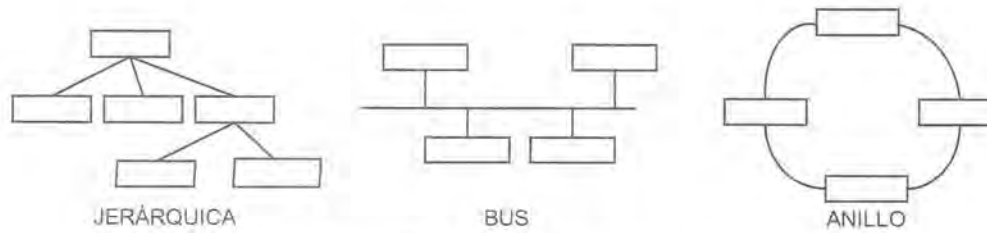


Figura 26.6. Topologías de redes.

Medios físicos

La interconexión de los nodos de una red se realiza mediante medios físicos muy diversos: cable, fibra óptica o enlace vía radio. La elección de uno u otro dependerá del tipo de aplicación y presupuesto económico, aunque debe tenerse presente que de su elección dependerá la velocidad de transmisión, la distancia máxima que podemos establecer entre los nodos y la fiabilidad del sistema (inmunidad a errores por interferencias, etc.). Lo más normal es la utilización de cable, aunque puede considerarse el uso de fibra óptica, por ejemplo, en el caso de entornos con nivel de ruido elevado (la fibra es muy inmune al ruido electromagnético) o en el caso de ambientes explosivos ya que los bajos niveles de energía que transporta la fibra previenen riesgos de explosión (seguridad intrínseca). No obstante, su coste suele ser algo mayor que el cableado clásico por lo que es importante su valoración (véase capítulo 20). También es muy común ver redes de control

donde se combinan las distintas alternativas adaptándolas a los requerimientos particulares que exige cada trazado. En cualquier caso, es necesario que la administración de la red se realice siempre de forma transparente independiente de su configuración física y sin disminución de prestaciones mediante la interfaz hardware necesaria (puentes, repetidores, etc.).

Control de acceso

La manera en que los nodos de la red acceden al medio físico para leer o recibir información y los mecanismos para controlar este acceso es otro aspecto fundamental en las redes y en especial en las redes de control. El objetivo es reducir el tiempo que tarda un nodo en acceder a la red para empezar a transmitir los datos (generalmente agrupados en paquetes de datos). Este factor es el factor fundamental para determinar, por ejemplo, si podemos considerar una red como válida para aplicaciones en tiempo real. Una red es apta para aplicaciones en tiempo real si la comunicación entre sus nodos permite dar respuesta a un acontecimiento (por ejemplo, la activación de una alarma) en un tiempo establecido y predecible. Las estrategias para asignar el medio físico más utilizadas son [26.3]:

- **Maestro fijo** (administración centralizada).
- **Maestro flotante** (administración distribuida).

Las estrategias del tipo maestro fijo dan lugar a las redes denominadas de **maestro-esclavo**, muy frecuentes como redes de control por tres razones: permite la implementación sencilla de ciertos aspectos jerárquicos de control, simplifica el funcionamiento de la red y abarata los costes de la interfaz física. En estas redes existe un nodo que toma el rol del controlador principal (o maestro) iniciando todas las transferencias entre los otros nodos (los esclavos) y pasando todas las transferencias por dicho nodo. Generalmente se utiliza un esquema de **encuesta cíclica** (*polling*) donde el maestro envía secuencialmente mensajes a los esclavos que responden por turno. Para el funcionamiento del *polling* es necesario que cada esclavo tenga su propia dirección. El maestro envía un mensaje comenzando con la dirección específica de un esclavo. El esclavo reconoce su dirección y ejecuta el comando incluido en el mensaje. Una vez hecho esto, se devuelve al maestro una confirmación y así podrá dirigirse al siguiente esclavo. El formato de la dirección y el mensaje son parte del **protocolo** que utiliza este sistema de control específico. El problema que se plantea es que cualquier fallo en este nodo da lugar a que se pare el sistema.

La estrategia de maestro flotante supone que no existe un nodo principal. El acceso se produce por parte de cada nodo sin intervención directa de otro. Así, la caída de un nodo no da al traste con todo el sistema. El problema que se plantea con esta estrategia es que debe resolverse cómo se accede al medio ya que no hay nadie que dicte el procedimiento de acceso. Las técnicas más usadas para controlar el acceso al medio son:

- **Paso de testigo** (*token passing*).
- **CSMA/CD**.

La técnica de paso de testigo parte de la existencia de un mensaje de tipo especial denominado "*token*" o testigo que circula por los distintos nodos de la red. Una estación puede tomar control del bus para transmitir sólo cuando tiene el control del testigo (el resto sólo puede recibir). Esta estrategia suele ser muy utilizada en los buses de campo, donde se implementa un anillo lógico con una topología física de bus pasando el testigo de estación en estación (*token-passing bus, IEEE-502.4*) (figura 26.7).

Resulta muy sencillo implementar turnos de ocupación del testigo basados en tiempos y establecer, de esta forma, prioridades. Así, aquellos nodos dedicados a procesos más críticos se les asignan tiempos mayores de ocupación de manera que es posible asegurar tiempos de respuesta mínimos. Este aspecto es muy útil en el control en tiempo real.

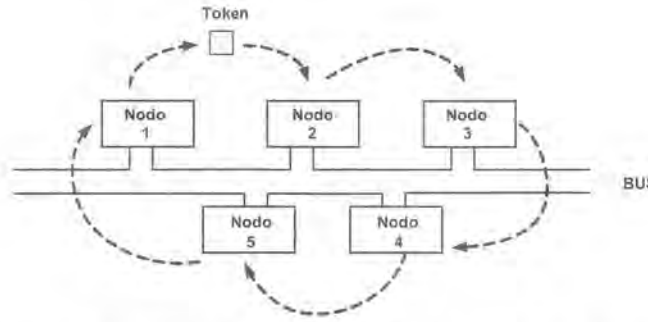


Figura 26.7. Topología *token bus*. La topología es de bus físico pero se define un anillo lógico por el que circula el testigo.

La técnica denominada CSMA/CD (*Carrier Sense Multiple Access/Colision Detected*) (IEEE-802.3) es muy utilizada en redes de datos como Ethernet y es apta para redes con topología en bus. El nodo que pretenda acceder a la red comprueba previamente si hay alguien ocupando el medio. Para ello, verifica la presencia de una portadora en la línea que situaría la estación que en ese momento estuviese transmitiendo (*Carrier Sense*). Si el medio se encuentra ocupado, se espera a que esté libre. No obstante, se plantea un problema cuando varias estaciones pretenden acceder simultáneamente al medio (*Multiple Access*); en ese caso, se detecta una colisión en el medio (*Colision Detected*) y las estaciones involucradas no vuelven a intentar acceder a este hasta después de pasado un intervalo de tiempo aleatorio. Este método plantea problemas cuando se pretende usar en un bus de campo ya que no garantiza un tiempo de respuesta determinado y por ello se desaconseja su uso en aplicaciones de tiempo real. Además, por las peculiaridades de las redes de control donde constantemente se están transmitiendo paquetes pequeños, pueden originarse múltiples colisiones que llevan a una degradación de las prestaciones del sistema.

En ocasiones pueden utilizarse métodos híbridos que combinan las características de paso de testigo y CSMA/CD.

Otro aspecto muy interesante en las redes se refiere a los procedimientos que existen para dirigirse a los nodos que componen la red (direccionamiento); se distinguen tres métodos:

- ***Unicast*** (punto a punto): el paquete se envía a una única estación.
- ***Multicast*** (a un grupo): el paquete se envía a un grupo de estaciones simultáneamente.
- ***Broadcast*** (difundido): el paquete se dirige a todas las estaciones de la red simultáneamente.

El direccionamiento de tipo *broadcast* es el más sencillo, sin embargo, puede suponer una sobrecarga para los nodos al tener que procesar información que no les afecta directamente. Los sistemas *unicast* y *multicast* son más eficientes y permiten establecer mecanismos que contribuyen a aumentar la fiabilidad como el acuse de recibo y el reenvío de paquetes. Este último aspecto es muy interesante en redes de control ya que en estas adquiere gran importancia el control de errores y su capacidad de recuperación [26.2].

26.3 Buses de campo

Los buses de campo son redes de control digitales multipunto, que conectan dispositivos de campo como controladores (PLCs), sensores y actuadores. Se corresponde con las redes de comunicación utilizadas en la

parte más baja de la pirámide mostrada en la figura 26.5 que es la parte más ligada al proceso, es decir, el nivel más bajo dentro de la escala de automatización.

Generalmente, los buses de campo facilitan las tareas de adquisición de datos, monitorización de señales, diagnósticos y gestión y supervisión de señales. Estas redes sustituyen a las antiguas redes de tipo analógico que utilizaban, generalmente, el enlace de 4-20 mA para la comunicación de las variables de proceso a un sistema central. La topología de red más utilizada en los buses de campo es de tipo bus y suelen configurarse como redes maestro/esclavo. Los elementos que componen un bus de campo son básicamente:

Maestros: *esta tarea la desempeña un controlador que generalmente es un computador o un PLC.* Normalmente existe un solo maestro por bus aunque pueden existir varios según la aplicación.

Esclavos: elementos pasivos o sin iniciativa propia de comunicación que suelen ser generalmente elementos actuadores o sensores.

Medio de conexión: cable o fibra generalmente.

Configurador de bus: se suele utilizar un computador y se encarga de establecer parámetros, asignar direcciones, mapear la información en el maestro, etc.

Supervisor: generalmente un computador que sólo toma información del bus.

Otros accesorios: utilizados para extender la capacidad del bus como por ejemplo: repetidores, *transceivers*, etc.

Las ventajas de los buses de campo pueden concretarse en:

- **Bus digital:** existe una ventaja inherente al utilizar un bus digital frente al tradicional sistema analógico tales como: menos interferencias, más precisión, más versatilidad, mayor simplicidad a la hora de obtener y transmitir los datos por parte de los dispositivos de campo que forman el bus, etc.
- **Inteligencia integrada:** la consideración de nodos inteligentes en el bus flexibiliza el sistema y permite que ciertos algoritmos de control se incluyan en el propio dispositivo de campo simplificando el sistema de control en su conjunto.
- **Flexibilidad y seguridad:** el esquema propuesto por los buses de campo se corresponde con un sistema de tipo distribuido, más seguro que uno centralizado, donde un fallo puede parar toda la instalación. Además, el control se realiza a través de sensores inteligentes y dispositivos de automatización que resuelven una parte del proceso de forma autónoma (islas de automatización) con capacidad de crecimiento y configuración muy diversas.
- **Coste:** representa una reducción de coste para el usuario final, fundamentalmente, por el ahorro en los costes de instalación, mantenimiento y otros derivados de la mejora del funcionamiento del sistema en general. La reducción en coste de cableado es notoria frente a una solución tradicional.
- **Estandarización:** el desarrollo de los buses de campo debe buscarse en la demanda de fabricantes e industrias de alcanzar un cierto grado de estandarización ante el crecimiento vertiginoso de los dispositivos de campo de distintos fabricantes. No obstante, el grado de estandarización no es tan alto como *en otros tipos de redes existiendo muchos buses propietarios. Aun así, muchas marcas han adoptado* ciertos estándares convirtiéndose en sus principales promotores.

Enlaces físicos de los buses de campo

En cuanto a las formas de enlace físico más utilizadas en las comunicaciones industriales se distinguen:

- RS-232-C (normas EIA-RS232, V.24 CCITT) (punto a punto 1 a 1).
- RS-422 (norma EIA-RS422, V.11 y X.27 CCITT) (multipunto 1 a n).
- RS-485 (norma EIA-RS485) (multipunto n a n).

Todos ellos son enlaces de tipo serie debido a la economía que supone frente una alternativa de comunicación en paralelo. En los buses de campo suele utilizarse el enlace RS-485 por sus prestaciones de inmunidad al ruido y distancias entre nodos. A continuación, analizaremos brevemente las características de cada uno de estos enlaces.

RS-232-C

Se trata de un viejo estándar de comunicación serie que se diseñó para la comunicación entre un equipo terminal de datos (DTE), como un computador y un equipo de comunicaciones de datos (DCE) como un módem. Es un enlace de tipo *full-duplex* y punto a punto (1 a 1). La distancia máxima sin ningún circuito de ampliación es de aproximadamente 15 m (en la práctica puede llegar a funcionar hasta con distancias de 100 m) y su velocidad típica ronda los 19.000 baudios aunque puede ser superior. Este tipo de enlace dispone de unas líneas de datos y unas líneas de control sobre las que se implementa el protocolo de comunicación (ciertas señales presentes en el conector que permiten el bloqueo de la transmisión o la recepción). La norma básica se ocupa de las especificaciones físicas del conector, los niveles de tensión de las señales y las señales de protocolo. Se suele utilizar un conector de 25 pines con formato DB25 (figura 26.8a). Cuando no son necesarias todas las señales se puede adoptar un conector de 9 pines (DB9) (figura 26.8b).

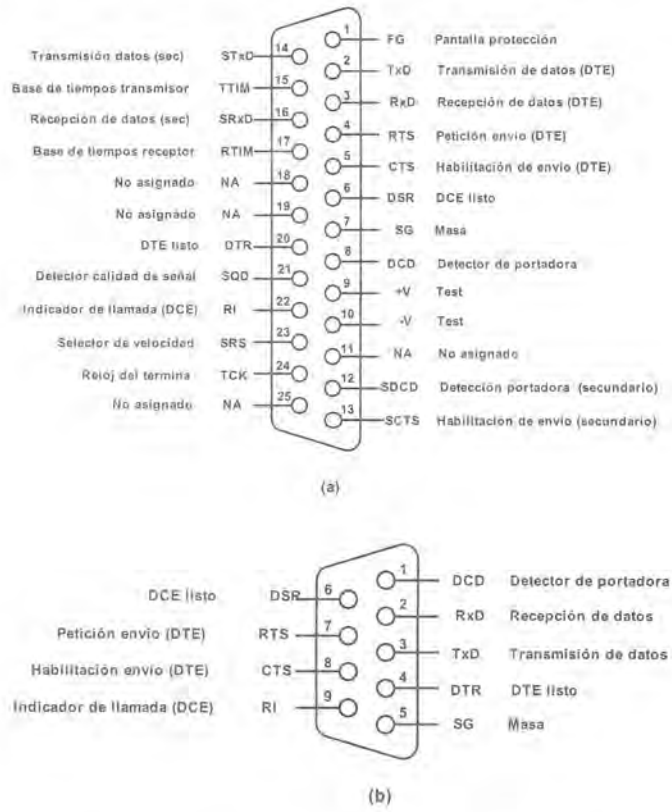

Transmisión datos (sec) STxD 14
Base de tiempos transmisor TTIM 15
Recepción de datos (sec) SRxD 16
Base de tiempos receptor RTIM 17
No asignado NA 18
No asignado NA 19
DTE listo DTR 20
Detector calidad de señal SQD 21
Indicador de llamada (DCE) RI 22
Selector de velocidad SRS 23
Reloj del termina TCK 24
No asignado NA 25
1 FG Pantalla protección
2 TxD Transmisión de datos (DTE)
3 RxD Recepción de datos (DTE)
4 RTS Petición envío (DTE)
5 CTS Habilitación de envío (DTE)
6 DSR DCE listo
7 SG Masa
8 DCD Detector de portadora
9 +V Test
10 -V Test
11 NA No asignado
12 SDCD Detección portadora (secundario)
13 SCTS Habilitación de envío (secundario)
(a)
DCE listo DSR 6
Petición envío (DTE) RTS 7
Habilitación envío (DTE) CTS 8
Indicador de llamada (DCE) RI 9
1 DCD Detector de portadora
2 RxD Recepción de datos
3 TxD Transmisión de datos
4 DTR DTE listo
5 SG Masa
(b)

Figura 26.8. Señales del conector RS-232 con (a) 25 pines y (b) 9 pines.

Los datos se transmiten por la línea TxD y se reciben por la línea RxD, en serie, es decir, bit a bit. Los niveles de tensión están referidos al terminal SG (línea de masa). Los niveles lógicos "0" y "1" se representan por los niveles físicos de tensión que se muestran en la figura 26.9:

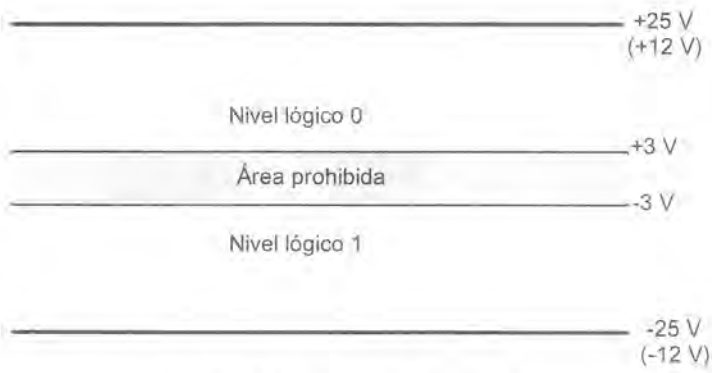


Figura 26.9. Niveles de tensión RS-232-C.

Cualquiera que sea el tipo de transmisión es necesario que el receptor se sincronice para saber en todo momento dónde comienza la transmisión de un bit, un carácter o un bloque. En la transmisión asíncrona, cada carácter va precedido de un bit de inicio (bit de *Start*) y finaliza con 1 ó 2 bits de parada (bits de stop) que garantizan la sincronización del receptor y permiten el reconocimiento del comienzo y el final del carácter. Esta es la forma de sincronización más simple (figura 26.10).

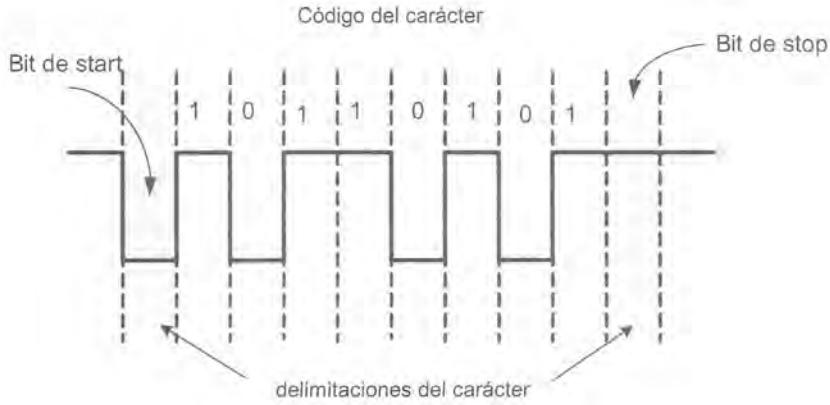


Figura 26.10. Transmisión asíncrona de un byte en serie.

Este tipo de enlace se utiliza para la conexión de una gran variedad de dispositivos con necesidades muy diversas por lo que a veces existe algo de confusión en torno a las señales que se utilizan y la forma en que se interpreta las señales de control de la interfaz. Para realizar la comunicación puede no ser necesario utilizar todas las líneas del conector, en especial las líneas que llevan la información de control. A veces, se utiliza exclusivamente las líneas TxD, RxD y SG sin ninguna otra señal de control por hardware. En este tipo de enlace el control del flujo de datos, en especial el inicio y fin de transmisión, se señala mediante unos caracteres especiales XON y XOFF, recibiendo la denominación de **enlace XON-XOFF**. La figura 26.11 muestra distintas posibilidades de interconexión conexión entre dispositivos según se considere como un DTE o un DCE.

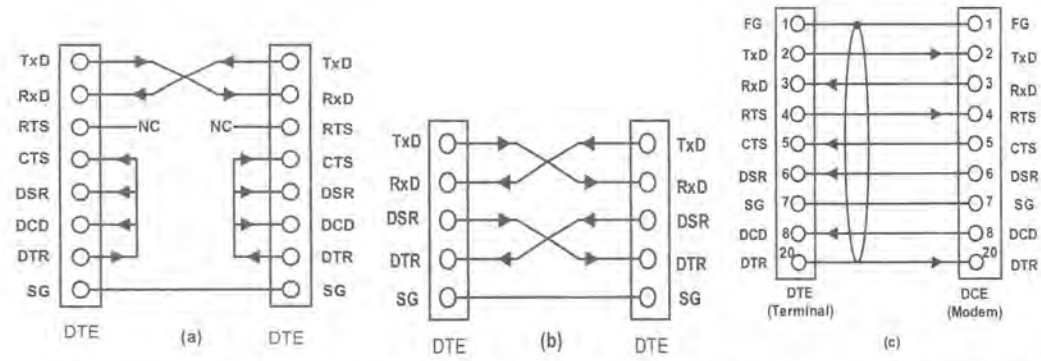


Figura 26.11. Distintas formas de enlace RS-232: (a) conexión serie entre 2 DTE sin señales de control, (b) conexión serie entre 2 DTE son señales de control, (c) conexión serie entre un DTE y un DCE.

La tabla 26.2 resume la función de las señales básicas de la interfaz RS-232, vistas desde un DTE.

Señal	I/O	Descripción
FG	–	(*Field Ground*) protección a tierra de los equipos.
TxD	O	Línea de transmisión de datos.
RxD	I	Línea de recepción de datos.
RTS	O	(*Request to Send*) El DTE está listo para enviar un dato. Se espera la respuesta a través de CTS.
CTS	I	(*Clear To Send*) Habilitación para transmitir. Generalmente la activa un receptor para indicarle al transmisor que está listo para recibir datos. Un dispositivo puede no transmitir si no está habilitada esta línea.
DSR	I	(*Data Set Ready*) Indica que el receptor está listo (encendido, conectado y preparado).
DTR	O	(*Data Terminal Ready*) El DTE está listo para transmitir o recibir (idéntica al DSR aunque desde un DTE).
DCD	I	(*Data Carrier Detect*) Indica que el DCE está dispuesto (ha recibido portadora). Para un DCE (módem) sería una línea de salida que indica que está recibiendo portadora. En algunos DTE se deshabilita la recepción si no se activa esta entrada.
SG	-	(*Signal Ground*) señal de referencia (masa).

Tabla 26.2. Descripción de las señales de la interfaz RS-232 desde el punto de vista de un DTE.

RS-422

Se trata de otro estándar de comunicación serie *full-duplex* que utiliza señales diferenciales. La figura 26.12 muestra la configuración de un enlace RS-422. Es una interfaz estándar muy apropiado para aplicaciones industriales que conectan un maestro con varios terminales. También permite la comunicación punto a punto entre dos nodos utilizando, generalmente, un par de cables trenzados para cada línea de señal (4 hilos). La transmisión en modo diferencial presenta como ventaja principal su inmunidad al ruido electromagnético (ruido de modo común). Los unos y ceros lógicos se establecen en función de la diferencia de tensión entre ambos conductores (figura 26.13). Permite establecer una mayor distancia de conexión, por lo cual es preferible frente a la interfaz serie RS-232, sobre todo, en entornos ruidosos como el industrial.

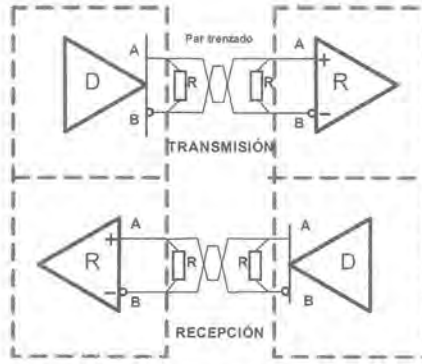


Figura 26.12. Enlace RS-422. Enlace punto a punto, *full-duplex*.

Las distancias y frecuencias que se utilizan en este bus son de 1200 a 1500 m a 100 Kbits/s o 50 m a 10 Mbits/s. Las líneas deben cargarse en los extremos con resistencia de terminación de línea (120 Ω generalmente). El propósito de la resistencia de terminación es prevenir la reflexión de los datos en el fin de la línea. Las señales de la interfaz se describen en la tabla 26.3.

Señal	Descripción
A	Entrada de recepción no-inversora y salida de emisión no-inversora
B	Entrada de recepción inversora y salida de emisión inversora

Tabla 26.3. Señales de un enlace RS422.

RS-485

Este enlace es uno de los más utilizados en la industria. Se trata de un enlace serie, con transmisión de señales en modo diferencial, multipunto. Este enlace está basado en la interfaz RS-422 pero utiliza sólo un par trenzado para la comunicación. Permite usar una topología de tipo bus, conectando los dispositivos en paralelo a los dos conductores, aunque desde un punto de vista lógico puede después organizarse como un anillo, estrella u otro tipo. Presenta una alta inmunidad al ruido y se pueden crear redes multipunto maestro-esclavo de forma muy sencilla (puede ver un ejemplo de implementación en el CD adjunto). La figura 26.13 muestra los niveles de tensión de la interfaz RS-485.

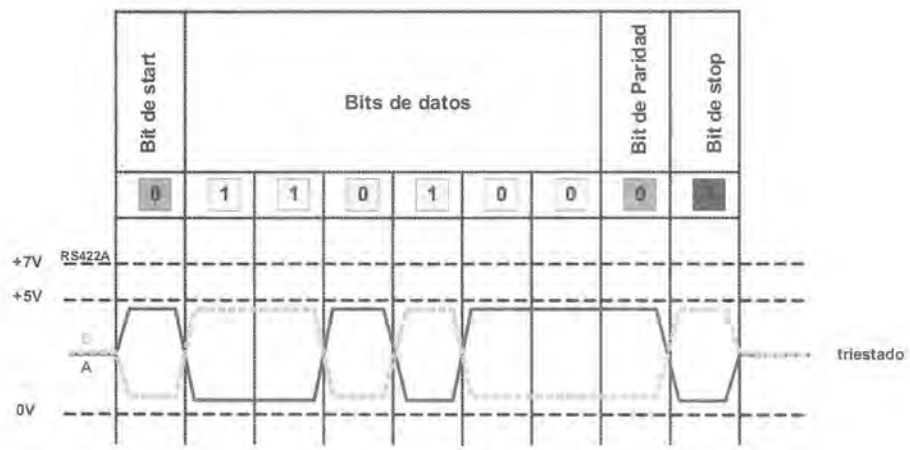


Figura 26.13. Niveles de tensión de las interfaces RS-485 y RS-422.

La figura 26.14 muestra un enlace punto a punto mediante RS-485 y en la figura 25.15 con una topología de bus. Normalmente, el número de nodos está limitado a unos 32 por razones de carga, pero puede ampliarse a más si se utilizan repetidores. Para aumentar la inmunidad al ruido se suele forzar el estado de la línea a un estado inactivo, si no existe un *driver* activo en la red, con un circuito formado por dos resistencias como se muestra en la estación 1 de la figura 26.15. Sin este circuito, es posible que la línea se vea afectada por ruidos que pueden activar los receptores y causar graves problemas en las comunicaciones. El control de flujo de datos se realiza generalmente mediante XON-XOFF. Las características del enlace en cuanto a niveles lógicos, distancia y velocidad de transmisión son idénticas al enlace RS-422.

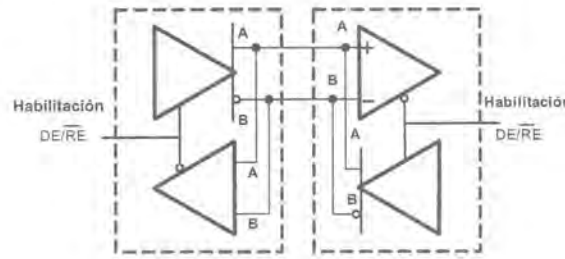


Figura 26.14. Enlace RS-485 punto a punto. La línea de habilitación DE/RE controla la transmisión (DE/RE = 1) y la recepción (DE/RE = 0) en cada estación.

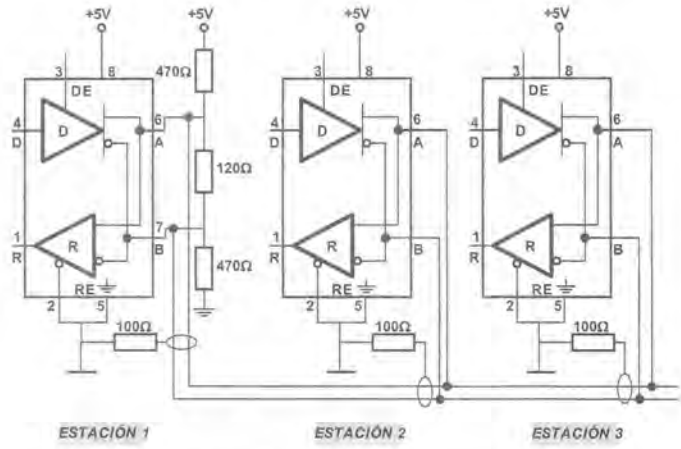


Figura 26.15. Enlace RS-485 en topología de bus.

El hecho de que exista una topología de tipo bus con varios dispositivos conectados implica la necesidad de administrar los accesos al bus ya que sólo un dispositivo debe de tomar control del bus para hablar en un determinado momento. La tabla 26.4 resume las características de los enlaces más utilizados.

	RS-232	RS-422	RS-485
Tipo de línea	Desbalanceada	Balanceada	Balanceada
Nº máximo de dispositivos	1	1	32
Nº máximo de receptores	1	32	32
Longitud máxima (m)	15	1200	1200
Velocidad máxima	20 Kb/s	10 Mb/s	10 Mb/s

Tabla 26.4. Comparativa de los enlaces serie RS-232, RS-422 y RS-485.

Buses de campo más importantes

Existen un gran número de buses de campo algunos de ellos propietarios debido, en parte, a la falta de estándares de facto. Cada protocolo está optimizado para diferentes niveles de automatización. En general podemos clasificar los buses de campo en tres grupos fundamentales [26.2]:

a) Buses de alta velocidad y baja funcionalidad: Diseñados para integrar dispositivos simples como finales de carrera, fotocélulas, relés y actuadores simples, funcionando en aplicaciones de tiempo real y agrupados en una pequeña zona de la planta, generalmente una máquina. Algunos ejemplos son:

- CAN (*Controller Area Network*): fue diseñado originalmente para su aplicación en el sector de automoción aunque debido a sus buenas características, es muy robusto y fácil de implementar, cada vez se usa más en la industria.
- AS-I (*Actuator Sensor-Interface*): bus serie diseñado por Siemens para la integración de sensores y actuadores. Es un sistema basado en transmisión de señales digitales paralelo/serie para reducir el cableado y los costes de instalación. Se trata de uno de los buses más extendidos en Europa.

b) Buses de alta velocidad y funcionalidad media: Estos buses son capaces de controlar dispositivos de campo complejos de forma eficiente y a bajo coste. En general, utilizan mensajes de tamaño medio permitiendo al dispositivo una mayor funcionalidad. Así, se incluyen aspectos como la configuración, la calibración o incluso la programación del dispositivo. Algunos ejemplos son:

- DeviceNet: está optimizado para el mercado de los dispositivos discretos (todo-nada) como detectores, actuadores e interruptores, donde el tiempo de respuesta y la repetibilidad son factores críticos. Utiliza como base el bus CAN.
- InterBus-S: bus de campo enfocado a la comunicación *full-duplex* entre sensores y actuadores.

c) Buses de altas prestaciones: Este tipo de buses gozan de las mayores prestaciones y soportan en general comunicaciones a nivel de toda la factoría. No obstante, debido a las elevadas prestaciones funcionales y de seguridad que soportan, pueden presentar algunos problemas de sobrecarga. Algunas de sus características son: posibilidad de configurar redes con varias estaciones maestras (redes-multimaestro), permitir comunicaciones maestro-esclavo, capacidad de direccionamiento *unicast, multicast y broadcast*.

Algunos ejemplos de este tipo de buses son:

- Profibus: fue uno de los primeros estándares de comunicación industrial y es uno de los más populares en Europa. La norma DIN 19245 recoge todas las especificaciones técnicas.
  - Profibus-FMS: orientado a la comunicación entre controladores (PLC) y dispositivos con velocidades medias (hasta 100 ms). Se corresponde con un bus de la parte media de la pirámide de la figura 26.3.
  - Profibus-DP: orientado a la comunicación del controlador con dispositivos de campo con velocidades medias (hasta 10 ms). Se corresponde con un bus de la parte baja de la pirámide. En la actualidad es el más extendido.
  - Profibus-PA: orientado a la automatización en la industria de procesos. A diferencia del DP, su capa física permite su uso en zonas peligrosas (seguridad intrínseca).
- Fieldbus Foundation.

En el CD se proporciona una tabla donde se comparan los principales buses de campo mencionados.

Resumen

En este capítulo se abordan algunos conceptos relacionados con las redes de comunicaciones en la industria centrándose en los buses de campo. Estos son redes de control digitales multipunto, que conectan dispositivos de campo como controladores (PLC), sensores y actuadores. Generalmente, los buses de campo facilitan las tareas de adquisición de datos, monitorización de señales y diagnóstico además de la gestión y supervisión de señales. Estas redes sustituyen a las antiguas redes de tipo analógico que utilizaban, generalmente, el enlace de 4-20 mA para la comunicación de las variables de proceso a un sistema central. La topología de red más utilizada es de tipo bus y suele configurarse como red maestro/esclavo. También se estudian los principales medios físicos de transmisión utilizados en los buses de campo, siendo el RS-485 uno de los más utilizados por sus características en cuanto a distancia e inmunidad al ruido.

Para finalizar se mencionan algunos de los buses de campo más extendidos. En general, se pueden clasificar como buses de campo de alta velocidad y baja funcionalidad (CAN, AS-I, etc); buses de alta velocidad y funcionalidad media (Device-Net, Interbus-S, etc.); y buses de altas prestaciones (Profibus, FieldBus Foundation, etc.).

27

Software de instrumentación

27.1 Instrumentos digitales
27.2 Software de instrumentación

27.1 Instrumentos digitales

Reciben el nombre de "instrumentos" los equipos que tienen como misión determinar la magnitud de una *variable, visualizarla, generarla, o convertirla en otra diferente. Los instrumentos electrónicos modernos son* en su mayoría digitales y pueden clasificarse en tres grandes grupos:

1. **Instrumentos de medida y visualización:** sistemas electrónicos que realizan la adquisición de una señal, su evaluación y la representación en forma gráfica o alfanumérica. El primer paso en este proceso es la adquisición de la señal para su posterior evaluación. La evaluación supone el procesado de la señal, de forma analógica o digital, para transformarla en información útil. La representación de datos permite una comunicación eficiente de los resultados al usuario. Podemos citar como ejemplos: multímetros, osciloscopios, analizadores de espectros, etc.
2. **Instrumentos generadores de señal:** son sistemas electrónicos que generan señales eléctricas de características establecidas, por ejemplo, los generadores de funciones.
3. **Instrumentos convertidores de señales:** son sistemas electrónicos que transforman una señal física cualquiera (eléctrica o no) en una señal eléctrica de características determinadas, en uno de cuyos parámetros está contenida la información de la variable física. Este tipo de instrumentos acondiciona la señal para que pueda ser medida utilizando los instrumentos del primer grupo descrito.

27.2 Software de instrumentación

Instrumentación virtual

Tradicionalmente, los instrumentos utilizados en los sistemas de medida se fabrican como dispositivos independientes con unas capacidades específicas de medida definidas por el fabricante. A este esquema respon-

den, por ejemplo, instrumentos como generadores de funciones, analizadores de espectros, multímetros y otros a los que hicimos referencia con anterioridad.

En la actualidad, un gran número de aplicaciones de medida y de automatización utilizan computadores y tarjetas de adquisición de datos (o instrumentos de adquisición) como etapa previa de captura de información, pero la funcionalidad exacta del sistema de medida, es decir, para qué sirve, se define por software aprovechando la capacidad de procesamiento del computador. El computador personal o PC juega aquí un papel clave por su coste, prestaciones y su elevada conectividad. La combinación de hardware en la tarea de adquisición de señal y de software de procesamiento, define un nuevo modelo de instrumento sobre una plataforma informática, donde la solución de instrumentación viene definida por el usuario y no por el fabricante. Una de las ventajas fundamentales de esta configuración es la flexibilidad. Así, puede concebirse un instrumento con la interfaz deseada, dotarle de la funcionalidad precisa y aprovechar la capacidad de visualización que brinda el computador. En este escenario, el software es el elemento clave; con el software adecuado, podemos procesar digitalmente los datos capturados por una tarjeta de adquisición y obtener, por ejemplo, su espectro de frecuencias, su valor eficaz o su valor máximo. En cierta medida, es como si contásemos con un analizador de espectros y con un multímetro físicamente, pero, en realidad, se prescinde de los instrumentos reales. Este concepto de instrumento se denomina "**instrumento virtual**" o "instrumento software" en contraposición con el "instrumento físico" o "instrumento real". En el primero su funcionalidad es dinámica y la define el usuario; en el segundo, su funcionalidad es estática y la define el fabricante.

LabVIEW

El éxito de la "instrumentación virtual" reside en la existencia de potentes herramientas que permitan un diseño sencillo y eficiente del software. Algunas gozan de gran popularidad como por ejemplo **LabVIEW**[1]. LabVIEW es un software especialmente concebido para la adquisición, el análisis y la representación de datos y está basado en un lenguaje de programación gráfico muy intuitivo. Los programas se componen de objetos conectados entre sí y organizados de forma jerárquica. Básicamente, el programador comienza diseñando la interfaz de usuario o apariencia del instrumento/programa[2] que pretende realizar: botones, interruptores, indicadores numéricos, gráficos, etc., para ello utiliza una ventana denominada panel frontal (figura 27.1) donde sitúa los objetos con los que interactuará el usuario. El programa en sí se confecciona en otra ventana denominada panel de control, donde enlaza y opera sobre los objetos que situó en el panel frontal (figura 27.2).

Figura 27.1. Panel frontal de un instrumento virtual correspondiente a un osciloscopio, pantalla donde se introduce el aspecto del instrumento.

[1] Laboratory Virtual Instrument Engeneering Workbench de National Instruments.

[2] En el contexto de LabVIEW los programas se denominan instrumentos virtuales.

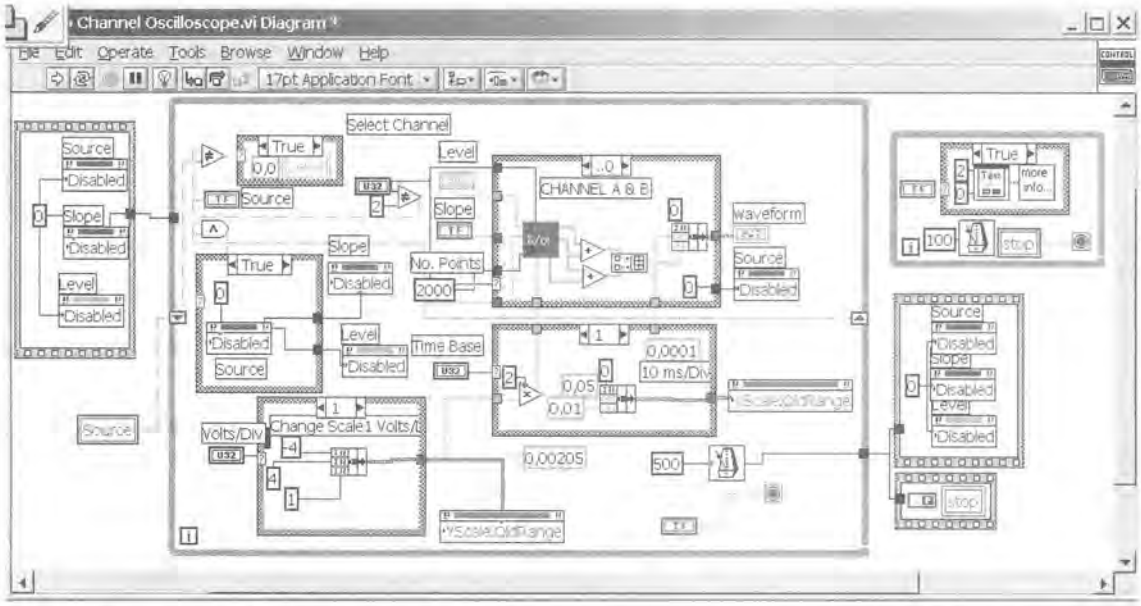


Figura 27.2. Panel de control del osciloscopio, pantalla donde se introduce el código.

Los objetos básicos que se utilizan en el panel de control son los iconos (que representa subprogramas), nodos (que realizan funciones e implementan estructuras de control) y terminales (que representan variables); todos estos elementos se enlazan por medio de cables, que se asimilan a tuberías por donde circulan los datos[3]. La programación se simplifica mucho gracias a que se dispone de una extensa librería de funciones para controlar de forma sencilla tarjetas de adquisición de datos y también multitud de instrumentos de test y medida. Generalmente, la comunicación con los instrumentos se realiza a través de interfaces como GPIB o interfaces serie, pero esta comunicación es totalmente transparente al usuario si utilizamos el módulo adecuado. LabVIEW también posee una extensa librería de funciones de procesamiento de señal y de representación de datos que puede completarse con librerías de funciones más específicas denominadas *toolbox* (librerías que contienen programas especializados). Se encuentran disponibles programas especiales para aplicaciones de tratamiento de imágenes, sonido y vibración, comunicación a través de Internet, procesamiento digital de señal, etc.

En la figura 27.1 se presenta un osciloscopio concebido como un instrumento virtual. El sistema está formado por una tarjeta de adquisición, que reside en interior del PC y el software de control adecuado. El instrumento funciona como si fuese un osciloscopio real pero su apariencia la especifica el usuario definiendo la interfaz más ajustada a sus necesidades y dotándole de la funcionalidad que precise.

Software SCADA (Supervisory Control And Data Acquisition)

SCADA se traduce como: Adquisición de Datos y Supervisión de Control. Todo proceso productivo con cierto grado de automatización debe disponer de un sistema de supervisión y control que proporcione la información imprescindible para la toma de decisiones basadas en la información procedente del proceso y de la organización. Un sistema SCADA es un software que permite la supervisión y el control de la producción. Mediante estos programas se puede convertir a un simple ordenador personal (PC) en un puesto de monitorización y control de la instalación capaz de comunicarse con los PLC de campo o dispositivos inteli-

[3] Puede ver una descripción detallada del lenguaje en la página Web del fabricante (www.ni.com).

gentes, permitiendo la representación gráfica de la planta, la supervisión del proceso, modificación de determinados parámetros, elaboración de informes, emisión de alarmas, etc. La información capturada es transmitida a un lugar conveniente (normalmente en una sala de control) y presentada de una manera compresible y utilizable. Las funciones básicas de un sistema SCADA son las que se describen a continuación:

- **Supervisión remota de instalaciones:** mediante esta función, una de las principales del sistema SCADA, el usuario puede conocer el estado de las instalaciones bajo su responsabilidad y coordinar eficientemente las labores de producción y mantenimiento en el campo.
- **Control remoto de instalaciones:** los sistemas SCADA permiten activar o desactivar equipos remotos (por ejemplo: interruptores, transformadores, bombas, válvulas, compresores, etc.) de manera automática o a solicitud del operador. Igualmente es posible realizar el ajuste de parámetros en los lazos de control analógicos y de comunicarse con dispositivos de campo o dispositivos de instrumentación.
- **Procesamiento de información:** en algunos casos, los datos capturados requieren procesamiento adicional a efectos de consolidar la información proveniente de diferentes lugares remotos.
- **Presentación de gráficos dinámicos:** se refiere a la posibilidad de visualizar los diagramas del proceso conteniendo información instantánea del comportamiento del mismo.
- **Generación de informes:** los sistemas SCADA permiten la generación automática o a petición de informes impresos de producción y balances.
- **Presentación de alarmas:** mediante esta función se alerta al operador sobre la ocurrencia de condiciones anormales o eventos que pudieran requerir su intervención. Normalmente, la importancia del evento o alarma se indica mediante el uso de colores y/o señales auditivas.
- **Almacenamiento de información histórica:** también permiten registrar y almacenar información operacional y de alarmas.
- **Presentación de gráficos de tendencias:** se pueden construir gráficos en tiempo real (o históricos) e inferir el comportamiento de las variables de proceso.
- **Programación de eventos:** se refiere a la posibilidad de programar la generación de informes, representación de diagramas del proceso o activación de tareas o comandos del sistema.

Otras funciones, no asociadas directamente a las operaciones en una sala de control, pero igualmente importantes son: diagnóstico de hardware y software, sistemas automatizados de mantenimiento preventivo y correctivo, integración con otros sistemas operacionales, integración con otros sistemas de información divisionales, calibración de equipos, etc.

El objetivo que se persigue con este tipo de sistemas es proporcionar un control y una supervisión integral del proceso productivo. Los programas SCADA suelen ser sistemas de arquitectura abierta, capaces de crecer y adaptarse a las necesidades de la empresa, de comunicarse fácilmente y de forma transparente al usuario y sencillos de instalar y programar. Es una herramienta indicada para directores o encargados de planta ya que permite controlar los errores, precisar dónde están y corregirlos gracias a la toma de datos *on-line*. Generalmente, un programa SCADA se compone de los siguientes módulos:

- **Módulo de configuración:** permite al usuario definir el entorno de trabajo de forma que se adapte a la aplicación particular que desea desarrollar.
- **Módulo de interfaz gráfico del operador:** proporciona al operador las funciones de control y supervisión de la planta. Generalmente, el proceso se representa mediante diagramas sinópticos que recogen la configuración de la planta (figura 27.4).

- **Módulo de proceso:** ejecuta las acciones de mando programadas teniendo en cuenta los valores actuales de las variables monitorizadas. La programación se realiza en lenguajes de alto nivel.
- **Módulo de gestión de archivos de datos:** se encarga del almacenamiento y el procesado de los datos permitiendo su acceso a través de otras aplicaciones o dispositivos.
- **Módulo de comunicaciones:** se encarga de la transferencia de información entre la planta y la arquitectura hardware que soporta el SCADA, y entre esta y el resto de elementos informáticos de gestión.

En cuanto a los componentes hardware que incluye están: el ordenador central **MTU** (*Master Terminal Unit*), los ordenadores remotos **RTU** (*Remote Terminal Units*), la red de comunicaciones y la instrumentación de campo. En la figura 27.3 se refleja la estructura básica de un sistema SCADA.

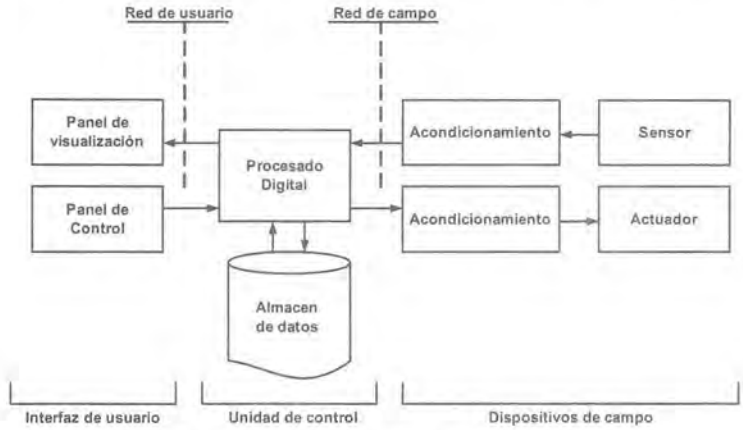


Figura 27.3. Esquema básico de un sistema SCADA.

Algunos ejemplos de programas SCADA o que incluyen software SCADA como parte de ellos son: Aimax, de Design Instruments, S.A., CUBE, Orsi España, S.A., FIX, de Intellution, Lookout, National Instruments, Monitor Pro, de Schneider Electric, SCADA InTouch, de LOGITEK, SYSMAC SCS, de Omron, Scatt Graph 5000, de ABB, WinCC, de Siemens.

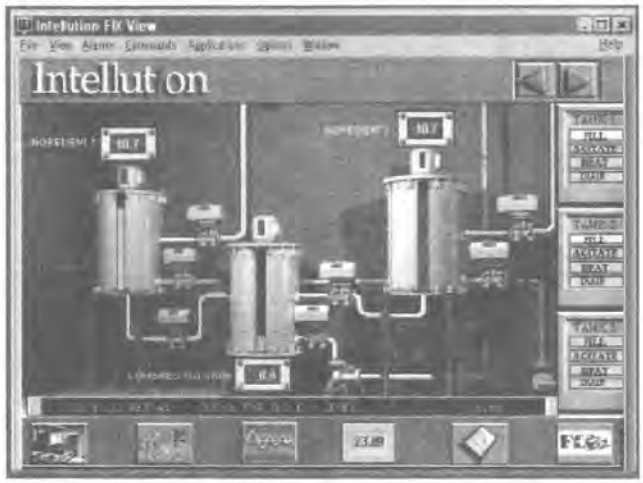


Figura 27.4. Ejemplo de una pantalla típica de un programa SCADA. La planta se representa gráficamente y se muestran los valores de diferentes variables del proceso.

Resumen

En este capítulo se introduce el concepto de instrumento virtual y se describen someramente algunas de las herramientas que gozan de una mayor popularidad para la programación de instrumentos virtuales. El concepto de instrumento virtual se contrapone al concepto de instrumento real. En un instrumento real, el típico equipo de medida o de generación de señal, la funcionalidad es fija y viene determinada por el fabricante, que establece las funciones básicas disponibles y la interfaz de usuario. En la instrumentación virtual la solución la define el usuario por medio del software. Para ello se utiliza generalmente una plataforma informática que actúa de soporte para el procesamiento de la información y para la representación de datos. La configuración más típica para construir un instrumento virtual de medida incluye una tarjeta de adquisición (interna en el PC) o un instrumento de adquisición externo, un computador (PC) y el software de control. Este último elemento juega un papel fundamental y por ello es necesario disponer de herramientas de desarrollo de software eficientes que estén especialmente adaptadas a tareas como el control de dispositivos de adquisición, el procesamiento digital de datos y la representación de la información. LabVIEW es un entorno de desarrollo basado en un lenguaje gráfico que facilita el diseño de aplicaciones de medida y automatización.

El capítulo concluye con el análisis somero de otro recurso software muy utilizado en el control, la automatización y la supervisión global de una planta denominada SCADA. Un programa SCADA realiza sus funciones en 4 niveles: Gestión (permite el intercambio de información para la toma de decisiones estratégicas en la factoría); Operación y supervisión (permite el mando y la adquisición de datos del proceso representando la planta de forma gráfica); Control (permite la comunicación con los dispositivos de control distribuido); Sensores y actuadores (permite la comunicación con los dispositivos de campo y los equipos de instrumentación).

28

Interferencias electromagnéticas

28.1 Introducción
28.2 Fuentes de interferencias
28.3 Acoplamiento de fuentes de interferencias
28.4 Minimización de los efectos de las interferencias

28.1 Introducción

En todos los sistemas electrónicos existen señales de cuya integridad puede depender el correcto funcionamiento del conjunto. La modificación de las señales puede producirse por muchos factores sobre los que influye el correcto diseño pero, desgraciadamente, no es fácil conseguir tener en cuenta todo lo necesario para garantizar ese correcto diseño: ya hemos visto en algunos capítulos anteriores la influencia de la distorsión o del ruido interno sobre los sistemas y hemos aprendido a reducir sus efectos o, cuando menos, a cuantificarlo y tenerlo en cuenta en la previsión de prestaciones del diseño. Sin embargo, una importante contribución a la modificación no deseada de las señales se debe a causas externas, mucho más difíciles de controlar o de predecir y que bien podrían dar al traste con cualquier diseño por muy correcto que este sea desde el resto de los puntos de vista.

28.1.1 DEFINICIONES

Todas estas influencias externas a una señal pueden incluirse dentro del término *interferencias*. A lo largo de este capítulo estudiaremos las causas de las interferencias y los procedimientos por los que se "cuelan" en nuestras señales y, en las dos siguientes, algunos de los remedios para conseguir que nuestros diseños no se vean afectados por ellas (en la medida de lo posible) ni sean fuente de problemas de este tipo para otros sistemas próximos.

Las interferencias tienen su origen fundamental en la presencia de campos electromagnéticos por lo que se ha acuñado el término EMI (*ElectroMagnetic Interferences* o interferencias electromagnéticas) para agru-

parlas y que se define como "el deterioro sobre la señal deseada inducido por el efecto de la radiación electromagnética generada por otro agente" [28.1]. Este deterioro puede ser más o menos grande y afectar más o menos a la operatividad de la señal, aspectos que no necesariamente tienen que ser proporcionales. Para comprenderlo, consideremos el ejemplo de una señal digital que sufre el efecto de una fuerte perturbación de su nivel de tensión por acción de una interferencia externa: siempre y cuando el nivel de esa perturbación no ponga en peligro el nivel lógico, el receptor de la señal interpretará la información correctamente y podría considerarse como nulo el efecto de la interferencia. Por el contrario, si la interferencia se produce sobre una señal analógica, el cambio del nivel de tensión puede afectar a la interpretación de la señal y, consecuentemente, perturbaciones pequeñas pueden llegar a ser muy molestas.

Como quiera que lo importante resulta preservar el comportamiento correcto del sistema, se ha definido un segundo término mucho más práctico que el primero que es el de Compatibilidad Electromagnética o EMC (*ElectroMagnetic Compatibility*) que queda definido como "medida de la tolerancia de un equipo a las perturbaciones electromagnéticas exteriores" [28.1] aunque puede particularizarse un poco más en la misma fuente:

a) Capacidad de los equipos y sistemas electrónicos para funcionar en un entorno de campos electromagnéticos a un determinado nivel de eficiencia.

b) La capacidad de un determinado componente de trabajar en su propio entorno electromagnético sin perturbaciones externas.

De las dos conveceiones, en los términos de este texto, haremos referencia únicamente a la primera por su mención explícita del comportamiento de los sistemas en entornos más o menos agresivos desde el punto de vista electromagnético.

Sin embargo, estas definiciones no son completas ya que plantean únicamente un papel de víctima al circuito, inmerso en un ambiente hostil cuando todo circuito puede ser susceptible de perturbar a otros y, en general, siempre se producirá la interacción de forma bidireccional. Por ello, una definición más completa de la compatibilidad electromagnética debe apuntar hacia **la capacidad de una instalación eléctrica o electrónica para trabajar satisfactoriamente en su entorno electromagnético y sin influir en él de forma no autorizada**.

En el caso particular de los sistemas de instrumentación, la presencia de señales muy débiles en las primeras etapas y la presencia de cableados más o menos complejos y de elevadas longitudes hace que sean sistemas potencialmente muy susceptibles a las interferencias, incluso en condiciones de pocas fuentes de problemas en las proximidades con lo que suelen asumir el papel de víctimas del entorno electromagnético.

Todas las señales inducidas sobre un sistema por la acción de campos electromagnéticos se sumarán a las generadas por las fuentes de ruido internas para obtener el total de ruido del sistema (figura 28.1); si en el capítulo 2 vimos cómo tener en cuenta el ruido interno y aprendimos a cuantificarlo en función de determinadas expresiones y datos de fabricantes, ahora veremos cómo tener en cuenta el externo aunque dado su carácter de no-controlabilidad, no podremos cuantificarlo.

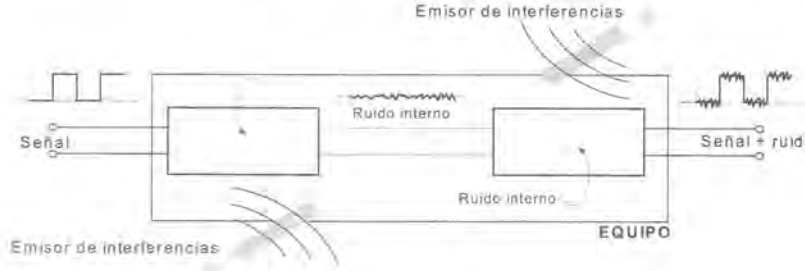


Figura 28.1. Causas de la aparición de ruido en un sistema.

Lo que sí podemos hacer es tratar de disminuir el efecto de las interferencias electromagnéticas de dos formas diferentes:

a) *Hacer que nuestro sistema o equipo sea capaz de soportar un ambiente electromagnéticamente adverso sin verse afectado por ello.*

b) Hacer que nuestro sistema no produzca emisión de interferencias electromagnéticas y pueda perturbarse a sí mismo o a otros equipos.

Para conseguir estos objetivos hay que prestar atención a las partes más susceptibles de actuar como receptores o emisores de interferencias, es decir, las líneas que transportan las señales por lo que nos centraremos fundamentalmente en los cables de conexión entre bloques o sistemas y en las pistas de circuito impreso que conectan los diversos componentes entre sí. Por ello, en el bloque constituido por el presente capítulo y los dos siguientes, trataremos en primer lugar las fuentes de interferencias y sus mecanismos de transmisión y, en segundo lugar, las formas de reducir los efectos tanto en el cableado e interconexión como en los circuitos impresos.

28.1.2 DESDE LA FUENTE A LA VÍCTIMA

En el proceso de acción de las interferencias hay que distinguir los diversos factores que influyen en él para poder actuar directamente sobre ellos y minimizar su negativo efecto sobre los circuitos. Para que se produzca un proceso de interferencia es necesario que exista una fuente de interferencia que emita la energía suficiente para perturbar el ambiente, un sistema víctima que sea susceptible a esa energía y un mecanismo de acoplamiento que la lleve desde la fuente a la víctima según se indica en la figura 28.2.

Figura 28.2. Acoplamiento del ruido entre el origen y la víctima.

Como cabe suponer, el objetivo final será la reducción del efecto del EMI y así, cualquier acción que tienda a disminuir la energía emitida por la fuente, la recibida por el circuito víctima o dificultar el proceso de acoplamiento entre ambos hará reducir el efecto de las interferencias. Veamos seguidamente qué fuentes y mecanismos de acoplamiento existen.

En general, como antes se ha indicado, todos los sistemas son, a la vez, fuente y víctima y considerar uno u otro sólo es una cuestión cuantitativa ya que los mecanismos de acoplamiento suelen ser bidireccionales; por la baja potencia que suelen manejar, los sistemas de instrumentación se consideran como víctimas en la mayoría de los casos. ¿Quiénes son, entonces, las fuentes?

La respuesta es sencilla: todos los sistemas eléctricos y electrónicos son fuentes potenciales de interferencias pero, además, existen fuentes naturales de energía que también pueden ser consideradas como tales. Pero, como cuestión cuantitativa que es, no todos los sistemas resultan perturbadores por lo que en la tabla 28.1 se indican algunos ejemplos de fuentes típicas de interferencias de tipo tanto natural como artificial [28.2], [28.3].

	Tipo	Ejemplo
Fuentes naturales de EMI	Terrestre	Rayos Relámpagos Descargas electrostáticas
	Extraterrestre	Energía solar Rayos cósmicos Ruido estelar
Fuentes artificiales de EMI	Sistemas eléctricos	Motores y generadores Líneas de distribución y transporte Aparallaje eléctrico Hornos de arco Equipos de soldadura Iluminación
	Sistemas electrónicos	Equipos de comunicación Radar Computadores y sistemas de control Fuentes de alimentación y accionadores Sistemas industriales
	Otros sistemas	Sistemas de encendido Sistemas de ultrasonidos Pulso nuclear

Tabla 28.1. Fuentes de interferencias.

La forma en que la energía procedente de estas fuentes termina afectando a nuestro sistema viene determinada por los mecanismos de acoplamiento que se pueden producir y que, básicamente son:

a) **Conducción:** la presencia de impedancias entre fuente y víctima o la existencia de caminos comunes de circulación de corriente en ambos produce este tipo de acoplamiento según se indica en las figuras 28.3a y 28.3c.

b) **Acoplamiento capacitivo o inductivo** en donde un conductor del sistema fuente y otro del sistema víctima actúan bien como placas de un condensador o como devanados de un transformador según se indica, respectivamente en las figuras 28.3b y 28.3d.

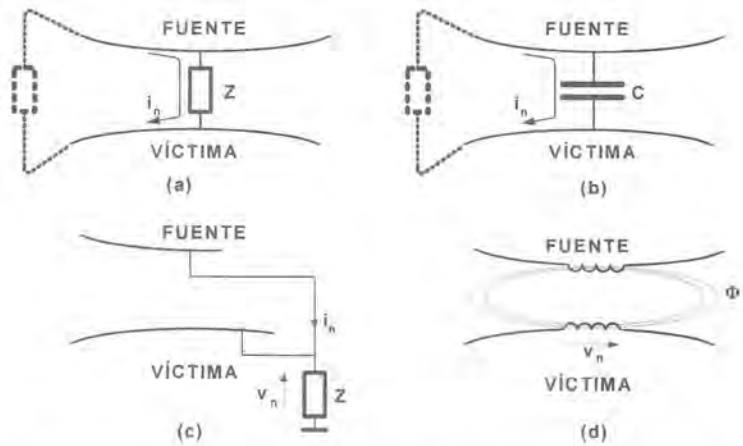


Figura 28.3. Diversos mecanismos de acoplamiento: (a) conductivo; (b) capacitivo; (c) por impedancia común; (d) inductivo.

Norma	Título	Fecha	Equivalencias
	Fuentes de generación de interferencias		
EN 60255	Relés eléctricos (Partes 22-6 y 25)	2001	IEC 60255
EN 60437	Ensayos de perturbaciones radioeléctricas de aisladores para alta tensión	1999	IEC 60437
EN 55022	Equipos de tecnología de la información. Características de las perturbaciones radioeléctricas. Límites y métodos de medida	2002	CISPR 22:1997
UNE 20589[1]	Fuentes de alimentación estabilizada de corriente continua. Ensayos de radiointerferencia	1979	IEC 478
UNE 20505	Límites y métodos de medida de las características de perturbación radioeléctrica de los vehículos a motor y dispositivos arrastrados por motores de encendido	1972	CISPR 12 :1978
UNE 20509	Características de las líneas de alta tensión relativas a las perturbaciones radioeléctricas. Descripción, del fenómeno	1999	CISPR 18 :1982
UNE 20511	Límites y métodos de medida de las características de perturbación radioeléctrica de los receptores de radiodifusión y equipos asociados	1993	CISPR 13-11:83
EN 12015	Compatibilidad electromagnética. Norma de familia de productos para ascensores, escaleras mecánicas y andenes móviles. Emisión	1998	
EN 60694	Estipulaciones comunes para las normas de aparamentas de alta tensión	1999	
EN 60495	Terminales de onda portadora sobre líneas eléctricas de alta tensión de banda lateral única	1996	IEC 495
EN 55013	Límites y métodos de medida de las características de perturbación radioeléctrica de los receptores de radiodifusión y equipos asociados	1999	
	Dispositivos para la eliminación de interferencias		
EN 133100 EN 133101 EN133200	Unidades de filtro pasivo para la eliminación de interferencias electromagnéticas	1998	
EN 60398	Inductores fijos para supresión de interferencias electromagnéticas	2000	IEC 60938
EN 132400 EN 132401	Condensadores fijos para la eliminación de interferencias electromagnéticas y conexión a la red principal	1996 1999	
UNE 208003	Recomendaciones para recintos apantallados	2002	R 210-005
	Compatibilidad electromagnética		
EN 61000	Compatibilidad electromagnética (CEM)	2002	IEC 61000
UNE 21000	Compatibilidad electromagnética (CEM)	1997	IEC/TR 1000
ENV 50141	Compatibilidad electromagnética. Norma básica de inmunidad	1996	
EN 60801 UNE 20801	Compatibilidad electromagnética para los equipos de control de procesos industriales	1996 1994	IEC 801
EN 60870	Equipos y sistemas de telecontrol (parte 2-1)	1997	IEC 870
EN 55103	Compatibilidad electromagnética. Norma de familia de productos para uso profesional de sonido, video, sistemas audiovisuales y para el control de iluminación de espectáculos	1997	
EN 55104	Compatibilidad electromagnética. Requisitos de inmunidad para aparatos electrodomésticos, herramientas y similares	1999	
EN 55014	Compatibilidad electromagnética. Requisitos de inmunidad para aparatos electrodomésticos, herramientas eléctricas y aparatos análogos	2002	CISPR 14-2:1997
EN 12016	Compatibilidad electromagnética. Norma de familia de productos para ascensores, escaleras mecánicas y andenes móviles. Inmunidad	1998	
EN 50091	Sistemas de alimentación ininterrumpida (SAI). Prescripciones para la compatibilidad electromagnética	1999	
UNE 26438	Vehículos automóviles. Perturbaciones eléctricas por conducción y por acoplamiento	1994	ISO 7637
ENV 50121	Aplicaciones ferroviarias. Compatibilidad electromagnética	2000	
EN 50083	Redes de distribución por cable para señales de televisión, señales de sonido y servicios interactivos (parte 8)	2002	
UNE 20725	Perturbaciones en los sistemas de radiocomunicación de los servicios móviles en presencia de ruido impulsivo	1993	CISPR 21
	Medida de las interferencias electromagnéticas		
UNE 208001	Especificaciones de los métodos y aparatos de medida de las perturbaciones radioeléctricas	1999	CISPR 16-1
EN 50147	Cámaras anecoicas	1997	

Tabla 28.2. Normas de AENOR relacionadas con las interferencias electromagnéticas.

[1] Esta norma será anulada próximamente por la nueva UNE-EN 61204.

c) **Radiación electromagnética** cuando el comportamiento de fuente y víctima se asemejan a antenas de emisión y recepción.

Como se puede ver en las distintas posibilidades de la figura 28.3 el efecto sobre el sistema víctima se traduce en una tensión o una corriente de ruido cuya importancia dependerá del nivel de señal que tenga el punto en donde se produce el acoplamiento por lo que los sistemas serán más sensibles a perturbaciones cuanto menores sean los niveles de señal presentes por lo que, una vez más será conveniente disponer lo antes posible de señales intensas para que los ruidos que hipotéticamente pudieran llegar ocasionen el menor efecto. Por tanto, habrá que cuidar especialmente las etapas de entrada de los sistemas de instrumentación y los sensores.

28.1.3 NORMATIVA

Se ha desarrollado una amplia normativa sobre los aspectos relacionados con las interferencias electromagnéticas y la compatibilidad electromagnética aunque, aun en la actualidad y como consecuencia del crecimiento de la presencia de sistemas electrónicos en todos los ámbitos de nuestra vida, se sigue trabajando en este aspecto. Si nos fijamos en la tabla 28.2, donde se recogen algunas de las normas de la AENOR relativas a la compatibilidad electromagnética, observaremos que muchas de ellas tienen fechas muy cercanas y otras han sido corregidas también muy recientemente.

El crecimiento exponencial de las comunicaciones de todo tipo, el mayor consumo energético y la presencia cada vez más abundante de equipos electrónicos de todo tipo en cualquiera de los ámbitos de la existencia hace que sea imposible predecir qué va a ocurrir en los próximos años en lo que a la EMC se refiere y algunas de las normas de la tabla 28.2 serán modificadas, derogadas o sustituidas por otras más modernas que tenderán a ser cada vez más restrictivas.

28.2 Fuentes de interferencias

En realidad, el término fuente de interferencia se usa para designar un lugar de origen de la energía que se transmite hacia otro lugar teóricamente no conectado con el primero y que se manifiesta en una tensión o corriente inducida. Desde este punto de vista, cualquier sistema, componente o camino de conexión es potencialmente una fuente de interferencia y lo único que necesita para manifestarse es la presencia de una pasarela para llevar la energía de uno a otro punto.

Sin embargo, algunos sistemas o componentes, por sus características específicas, son más proclives a actuar como fuentes de interferencias ya sea por la potencia que están manejando o por la facilidad para ponerla "a disposición" de otros sistemas. En la tabla 28.1 se han mostrado algunas de las fuentes de interferencias más típicas con las que nos podemos encontrar, aunque ahora vamos a efectuar una clasificación un poco distinta siguiendo el principio de cómo se origina la energía interferente y de cómo se manifiesta en el ambiente:

a) Sistemas y componentes sin transitorios (cableado de equipos, tarjetas o sistemas, cableado de red eléctrica, transformadores, bobinas, motores de jaula de ardilla).

b) Sistemas y componentes con transitorios (sistemas conmutados electrónicamente tales como sistemas digitales de alta frecuencia y equipos de potencia).

c) *Sistemas y componentes susceptibles de generar arcos (componentes electromecánicos de contacto* tales como relés o contactores, transformadores con tomas, motores de continua y de colector, hornos de arco).

d) Sistemas de transmisión de radiofrecuencia (emisoras de radio, TV, telefonía celular e inalámbrica, microondas, etc.).

e) Otras fuentes (sistemas mecánicos, descargas atmosféricas, explosiones nucleares, etc.).

Seguidamente se estudiarán someramente algunas de las fuentes más habituales con las que podremos encontrarnos y que afectarían a cualquier diseño de un equipo electrónico para instrumentación. Lógicamente, estudios más detallados quedan fuera del alcance de un libro general de instrumentación pero pueden ser ampliados en las fuentes bibliográficas indicadas [28.4].

28.2.1 SISTEMAS Y COMPONENTES SIN TRANSITORIOS

La mera circulación de corriente por los conductores o la presencia de tensiones más o menos elevadas en cualquier punto de un sistema eléctrico o electrónico dan lugar a la aparición de campos eléctricos y/o magnéticos capaces de generar señales sobre cualquier otro conductor víctima presente en esos campos. Veamos algunas de estas fuentes generadoras de campos electromagnéticos.

Máquinas eléctricas sin contactos mecánicos

Dentro de este epígrafe trataremos las máquinas eléctricas estáticas (transformadores y bobinas) y algunas máquinas rotativas como los motores de jaula de ardilla que son capaces de producir perturbaciones.

Los transformadores están constituidos por dos o más arrollamientos sobre un núcleo magnético de alta permeabilidad; en teoría, el campo generado por uno de los arrollamientos (el primario) se encamina por el núcleo magnético e induce tensiones en los demás (secundarios) tal y como se muestra en la figura 28.4a en que se muestra un transformador con un primario y un secundario.

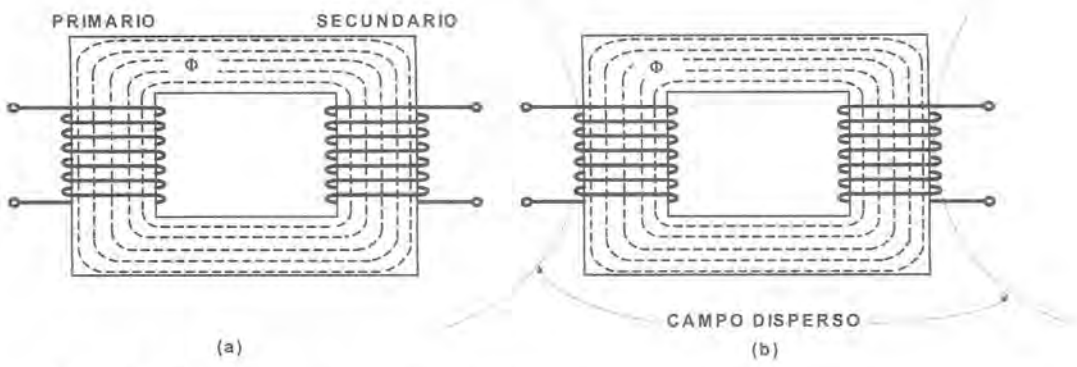


Figura 28.4. Transformadores como fuentes de interferencias: (a) comportamiento ideal; (b) el campo disperso se cierra por el aire y es capaz de producir interferencias.

En la práctica, el campo magnético no sólo se cierra por el núcleo, sino que también lo hace por el aire (campo disperso que se modela como una inductancia de dispersión [28.5]) con lo que ese campo puede afectar a los sistemas eléctricos y electrónicos de las proximidades. Obviamente, este efecto resulta mucho

más intenso cuanto mayor sea la potencia que maneja el transformador y menor la permeabilidad magnética del núcleo[2].

Las bobinas presentan el mismo problema intrínseco que los transformadores desde el punto de vista de la generación de interferencias aunque carecen de secundario: el campo magnético que se cierra por el aire es susceptible de producir problemas en otros sistemas cercanos. En el caso de las bobinas suele ser frecuente la presencia de entrehierro (*air gap*) con lo que los efectos mencionados se ven incrementados por el incremento de la reluctancia del camino magnético a través del núcleo y por la dispersión generada en el propio entrehierro.

En el caso de los motores de inducción de jaula de ardilla, el campo magnético está teóricamente confinado en el entrehierro que separa el rotor del estator pero, en la práctica tampoco es así del todo y una parte del flujo se cierra por caminos externos contribuyendo a generar campo perturbador en las proximidades del motor.

A pesar de que cualquiera de las máquinas eléctricas tanto estáticas como rotativas son fuentes potenciales de interferencias, un diseño cuidadoso en este aspecto las reduce mucho; en general, suelen estar bastante bien diseñadas, sobre todo en cuanto las potencias sean grandes o las frecuencias elevadas ya que, como sistemas de gestión de energía que son, deben procurar transferir la energía de un lado al otro y no dispersarla (pérdida de rendimiento). Sin embargo, hay que tener cuidado con transformadores de red de pequeña potencia y bajo coste como los que suelen acompañar a algunas fuentes de alimentación lineales cuyo diseño no está demasiado cuidado y que pueden ser fuente de una perturbación constante de 50 Hz que aparece en todo el equipo.

Conductores eléctricos en general

Cualquier conductor por el que circule una corriente es capaz de producir un campo electromagnético a su alrededor determinado por las leyes de la física [28.6] y que depende, en general, de la forma geométrica del conductor y de la corriente (figura 28.5). Si la corriente es variable, el campo electromagnético resulta también variable y es capaz de inducir señales en sistemas eléctricos y electrónicos próximos de mayor valor en la medida en que sea mayor la frecuencia de variación.

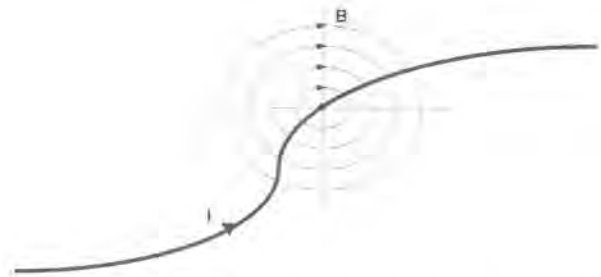


Figura 28.5. Campo magnético originado por un conductor por el que circula una corriente I.

De la misma forma, la presencia de una tensión elevada en un determinado conductor da lugar a la generación de un campo eléctrico (figura 28.6) en sus proximidades. Como quiera que la presencia de tensiones y

[2] Si la diferencia de permeabilidad magnética del núcleo y del aire es muy grande, el campo tendrá tendencia a cerrarse por el núcleo (camino de menor reluctancia) y disminuirá el campo disperso, mientras que cuando la permeabilidad del núcleo es sólo algo mayor que la del aire, el campo se cierra por el aire en bastante medida y aumenta el campo disperso. Este es el caso de los núcleos de ferrita en aplicaciones de potencia en alta frecuencia.

corrientes suele ser simultánea en líneas generales, lo habitual es tener ambos efectos en las proximidades de cualquier conductor.

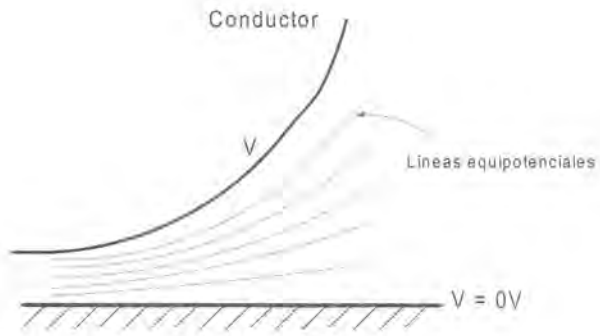


Figura 28.6. Producción de campo eléctrico por un conductor sometido a una tensión V.

Teniendo estos aspectos en cuenta, cualquier conductor por el que circulen corrientes variables elevadas o que esté sometido a tensiones elevadas tenderá a ser una fuente de EMI. Esto ocurre en líneas de transmisión de potencia eléctrica que llegan a mostrar sus efectos a decenas de metros de distancia, en cableado por el que circulen señales de alta frecuencia y en pistas de circuito impreso en sistemas de alta frecuencia o que manejen tensiones relativamente altas [28.7].

El efecto de los campos electromagnéticos cae rápidamente con la distancia a la fuente por lo que las precauciones tendrán que extremarse cuando las fuentes estén próximas a las posibles víctimas ya que, en esos casos, incluso señales no muy intensas pueden inducir interferencias en circuitos vecinos.

28.2.2 SISTEMAS Y COMPONENTES CON TRANSITORIOS

Los transitorios que se generan en el funcionamiento de los sistemas y que aparecen en la conexión y desconexión de etapas, sistemas, cargas, etc. implican el corte o aplicación de corrientes, lo que genera o puede generar oscilaciones transitorias [28.8]. Esto supone que estaremos en una situación similar a la del caso anterior pero con frecuencias determinadas no por la frecuencia de trabajo del sistema sino por los valores de las capacidades e inductancias que se conecten o desconecten. Esas corrientes y tensiones oscilatorias pueden tener frecuencias bastante mayores que la de trabajo y contribuir en gran medida a la generación de interferencias cuando circulan por cables de conexión o por pistas de circuitos impresos.

Claro está que la presencia de oscilaciones transitorias es más grave en la medida que la potencia que se esté aplicando sea mayor por lo que resulta de especial importancia como fuente de generación de interferencias los casos de equipos de electrónica de potencia:

a) *Cuando se efectúan conexiones o desconexiones de cargas de potencia con equivalentes RLC* mediante cualquier tipo de conmutador (estático o electromecánico) tanto en los casos de alterna como en los de continua.

b) Los equipos electrónicos de potencia que trabajan en conmutación (la mayoría de ellos). Durante su funcionamiento se efectúan muchas operaciones de corte y conexión de determinadas etapas del circuito lo que genera un buen número de transitorios. Resulta de especial importancia aquellos sistemas que trabajen en alta frecuencia (como los convertidores DC/DC y las fuentes de alimentación) y aquellos que trabajen con corrientes y tensiones elevadas, como es el caso de los inversores, algunos tipos de rectificadores controlados y convertidores AC/AC aunque su frecuencia sea menor.

c) Los sistemas digitales de alta frecuencia[3] se caracterizan por conmutar circuitos y cargas de baja o muy baja potencia pero la presencia de parásitos en las líneas (resistivos, inductivos y capacitivos), así como la consideración de las conexiones como líneas de transmisión —potenciales generadoras de rebotes [28.9]— hace que se produzcan abundantes fenómenos transitorios susceptibles de generar interferencias.

d) Circuitos de mando de relés y contactores. Este tipo de accionamientos muy comunes en el mundo de la electrónica industrial son dispositivos que permiten el cierre y apertura de circuitos mediante un electroimán sobre el que actúa el sistema de control. Este electroimán posee una bobina sobre la que se actúa para producir el cierre o apertura del dispositivo haciendo circular una corriente o cortándola. Tanto en un caso como en otro pero, sobre todo, en el segundo, se ocasiona un transitorio bastante intenso con sobreoscilaciones capaces de producir fuertes interferencias. En la figura 28.7 se proporciona un ejemplo de un sistema de este tipo.

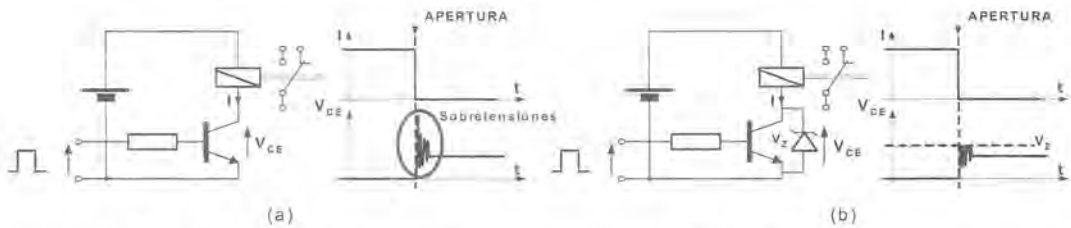


Figura 28.7. Efecto de la activación/desactivación de las bobinas de relés:
(a) producción de sobretensiones en la apertura; (b) limitación de las sobretensiones mediante zéner.

El problema se puede resolver parcialmente usando dispositivos de absorción de sobretensiones (VSD *Voltage Surge Devices*) o diodos zéner o de avalancha según se indica en la figura 28.7b. En el caso de relés o contactores gobernados en alterna, se pueden usar dos de los dispositivos de protección (VSDs o zéners) en antiserie para paliar el problema.

Claro está que la mejora de los diseños de cualquiera de los circuitos mencionados puede paliar los problemas de interferencias en la mayoría de los casos: por ejemplo, la adaptación de impedancias en las conexiones que deban ser consideradas como líneas de transmisión, la disminución de los parásitos, la utilización de redes de amortiguamiento (*snubbers*) en los equipos de potencia, etc.

28.2.3 SISTEMAS GENERADORES DE ARCOS

Durante los procesos de desconexión de cargas con un elevado componente inductivo la presencia de una corriente elevada puede provocar la aparición de un arco que trate de mantener el valor de la corriente. El arco de apertura de un circuito producido entre dos contactos es un proceso de descarga eléctrica en cualquier medio ocasionado por la elevación de la tensión entre los contactos por encima del máximo que se puede soportar.

Arcos de apertura

Sea por ejemplo un circuito como el de la figura 28.8 que tiene el interruptor S cerrado y se está aplicando la fuente sobre la carga RL.

[3] El concepto de "alta frecuencia" es una idea relativa y no se está hablando del mismo valor de frecuencia cuando se trata de convertidores DC/DC que cuando se trata de sistemas digitales. En el primera caso, se puede considerar que un convertidor es de alta frecuencia cuando supera los 100 kHz de frecuencia de conmutación mientras que esa es una frecuencia irrisoria para un sistema digital que podría llegar a tener frecuencias de reloj próximas a 1 GHz.

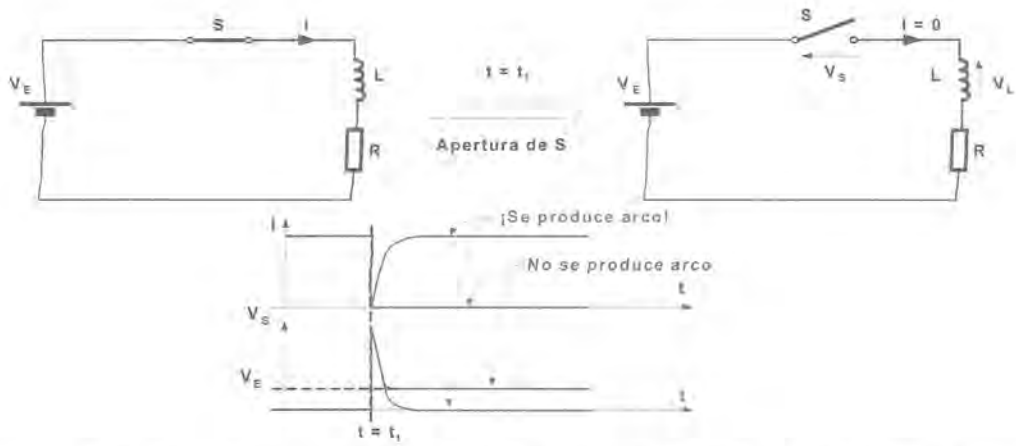


Figura 28.8. Posible producción de un arco en un contacto cuando se abre una carga RL.

En el instante t_1 se abre el interruptor S (puede ser un contacto de un relé o algo similar) con lo que la corriente por la carga se interrumpe bruscamente; como quiera que una variación brusca de corriente en la inductancia ocasiona una tensión muy importante, el interruptor S después de la apertura se ve sometido a una tensión muy alta. Si la máxima capacidad de soportar tensión del interruptor es sobrepasada, saltará el arco entre los contactos, restableciéndose así la circulación de corriente (en el fondo, la inductancia trató de mantener el valor de la corriente).

La máxima tensión que se soporta depende del gas o fluido entre contactos, de la temperatura, de la presión y de las características geométricas del interruptor (forma y dimensiones de los contactos y separación entre ellos). Si se supera este valor, el arco saltará uniendo los contactos con una caída de tensión baja y el circuito quedará nuevamente cerrado.

La presencia de cargas inductivas es muy frecuente en circuitos electrónicos de potencia pero, aun en el caso de no existir explícitamente como tales, los parásitos inductivos del cableado y conexionado y los de los diversos componentes, actúan de la misma forma que antes se ha explicado.

También se puede generar un arco en los fusibles durante el proceso de fusión por sobrecorriente [28.10]: al quemarse el fusible, se abre un circuito y cualquier componente inductiva tenderá a mantener la corriente con lo que el fenómeno antes descrito se produciría en el propio fusible.

La presencia de arcos es una fuente de interferencias muy importante y puede ocasionar severas molestias a los circuitos de las cercanías tanto durante el proceso de producción como durante el tiempo en que se mantenga debido a sus oscilaciones.

Hay muchos sistemas susceptibles de provocar arcos pero serán más importantes aquellos que manejen mayores niveles de potencia por lo que habrá que prestar especial atención a:

a) Equipos de potencia con interruptores electromecánicos (relés, contactores, etc.).
b) Sistemas de protección de sobrecorriente en equipos de potencia o en instalaciones eléctricas (fusibles, relés magnetotérmicos, etc.) o sistemas de absorción de sobretensiones.
c) Sistemas para los que el arco sea un objetivo de funcionamiento tales como equipos de iluminación mediante lámparas de arco propiamente dichas (Xe, H_2 o N_2) o lámparas de alta intensidad de descarga.

Arcos de cierre y rebotes

También es posible la aparición de arcos durante los procesos de cierre de los interruptores como consecuencia del proceso de acercamiento de los contactos unos instantes antes del cierre: como antes se indicó,

un arco se produce cuando se supera la rigidez dieléctrica del medio que es función (entre otras cosas) de la distancia entre contactos; pues bien, cuando un interruptor se va a cerrar, la distancia es muy pequeña y puede que la tensión aplicada entre contactos supere ese valor de máxima tensión impuesta por la rigidez dieléctrica. En ese caso se produce un pequeño arco de cierre que se extingue en el momento en que los contactos se cierran definitivamente; su importancia es menor cuantitativamente pero su presencia en cualquier entorno y de forma independiente de las cargas inductivas hace que sean poco previsibles por lo que pueden constituir una fuente de interferencias difícil de localizar.

Los rebotes mecánicos durante los procesos de cierre de contactos pueden originar tanto arcos de apertura como arcos de cierre: son procesos cierre-apertura-cierre repetidos por el choque elástico entre los contactos de un interruptor durante el proceso global de cierre que ocasiona conexiones y desconexiones en distancias contacto-contacto muy pequeñas por lo que pueden saltar microarcos que se extinguen y se reinician muchas veces con lo que contribuyen a la generación de interferencias. El fenómeno de los rebotes se suele tratar en los sistemas digitales para hacer que no afecten al comportamiento del circuito ni a su funcionalidad pero este tratamiento no suele suponer su eliminación sino sólo la insensibilidad de la entrada a la que está conectado el interruptor ante la presencia de rebotes. Sin embargo, aunque el sistema no perciba esos rebotes, sí que seguirán estando y serán la causa de interferencias electromagnéticas. Para su completa eliminación habrá que usar buenos interruptores que minimicen los errores mediante un correcto diseño mecánico de su sistema de cierre.

28.2.4 OTRAS FUENTES DE INTERFERENCIAS

Sistemas de transmisión de radiofrecuencia

Aunque el mecanismo de transmisión de señales mediante radiación queda fuera del alcance de este libro, no cabe duda que la presencia de generadores de radiación ocasiona campos electromagnéticos importantes, incluso a distancias grandes que pueden afectar a los circuitos generando tensiones y corrientes. No obstante, si la banda de frecuencias de trabajo de estos está muy alejada de las frecuencias a las que los circuitos víctima están funcionando resultarán poco relevantes; este suele ser el caso de la mayoría de los circuitos de instrumentación electrónica —excepto aquellos que se dedican a la medida de radiofrecuencia en particular— que se suelen manejar ya que las frecuencias de transmisión de radiofrecuencia se sitúan en valores muy altos y alejados de los anchos de banda de los circuitos de instrumentación.

No obstante, en las proximidades de los transmisores pueden inducirse niveles suficientes como para saturar alguna de las etapas ocasionando un mal funcionamiento del conjunto aunque las señales interferentes queden fuera del ancho de banda del sistema víctima. También hay que tener especiales precauciones cuando se empleen sistemas de telemedida con enlace por ondas hertzianas ya que los sistemas receptores podrían verse perturbados en sus primeras etapas aunque las frecuencias de transmisión sean bien distintas de las interferentes.

Descargas atmosféricas

Las descargas atmosféricas liberan una gran cantidad de energía que es una importante fuente de interferencias que afecta a la totalidad de los equipos eléctricos y electrónicos; es muy difícil que la descarga afecte directamente a un equipo pero la mayoría se ven afectados por la conducción de interferencias a través de la red eléctrica (esta sí que se ve afectada directamente con frecuencia). Dada la energía que se pone en juego en este caso, las interferencias pueden llegar a tener carácter destructivo por lo que las protecciones se encaminarán fundamentalmente a salvar al equipo de la incidencia.

Lógicamente, dado lo difícil de prever tanto la aparición de este tipo de fenómenos como su magnitud ya que dependen de circunstancias muy diversas, la mayoría de los sistemas de instrumentación no suelen estar

protegidos especialmente contra estas incidencias, sino que se limitarán a un adecuado sobredimensionamiento de los bloques más susceptibles a sufrir perturbaciones de la red eléctrica: sus fuentes de alimentación.

Pulso nuclear

Aunque afortunadamente las explosiones nucleares no son algo frecuente y esperemos que en el futuro se mantenga esta idea, algunos de los equipos y sistemas instrumentales tienen que ser capaces de resistir el efecto del denominado pulso nuclear (NEMP o *Nuclear ElectroMagnetic Pulse*) sobre todo aquellos sistemas de carácter militar o los que deban seguir en funcionamiento después de un suceso de este tipo.

El NEMP es una onda electromagnética originada por dos efectos: la ionización del aire y la liberación de partículas β (electrones) entre otras con una duración muy corta —unos pocos microsegundos— y un tiempo de subida al máximo nivel de señal de pocos nanosegundos. Su efecto es el de un previsible daño en los circuitos víctima provocados por altos niveles de energía electromagnética por lo que la protección necesaria sería similar a la utilizada contra las descargas atmosféricas.

Interferencias de origen mecánico

La generación de tensiones en un determinado circuito ante la presencia de un campo magnético sólo se produce cuando el campo es variable. Para ello, es posible que la fuente genere un campo variable provocado por una corriente, variable a su vez, o que el cable se mueva cuando la corriente es constante; esto es algo muy frecuente en máquinas sometidas a vibraciones en donde los cables eléctricos no estén firmemente sujetos a la estructura o carrocería de la máquina; aun en presencia de corrientes continuas, los movimientos de los cables producidos por la vibración dan lugar a interferencias de frecuencia igual a la del movimiento mecánico.

Otro fenómeno que puede originar interferencias es el efecto triboeléctrico que consiste en la producción de cargas libres por rozamiento o deformación de un conductor respecto su aislante y que variará al estar sometido a vibraciones o movimientos; este efecto puede ser notorio en los cables de los sensores que manejan señales débiles como en el caso que se comentó en el capítulo dedicado a los sensores piezoeléctricos.

28.3 Acoplamiento de fuentes de interferencias

Para que una interferencia generada por una fuente termine por afectar a un circuito víctima es imprescindible que exista un mecanismo de acoplamiento que actúe como pasarela para que la energía interferente generada por la fuente acabe llegando a la víctima.

Ya se introdujeron las posibles formas de acoplamiento fuente-víctima y ahora las trataremos con un poco de detalle para comprender la forma en que se producen y, así, evitar que se manifiesten con lo que habremos logrado que nuestros circuitos sean un poco más inmunes a las interferencias y resulten más compatibles con los ambientes hostiles.

Básicamente, tenemos dos formas de acoplamiento fuente-víctima: cuando existe un camino para la circulación de corriente entre ambos (acoplamiento conductivo) o cuando no existe (acoplamiento capacitivo, inductivo o por radiación); seguidamente estudiaremos todas estas formas de acoplamiento.

28.3.1 ACOPLAMIENTOS CONDUCTIVOS

Este tipo de acoplamiento se produce cuando hay un camino —explícito o no— para la circulación de corriente entre la fuente y la víctima o cuando ambas comparten una parte de los caminos de circulación de

corriente. Se pueden dar varios casos para que aparezcan impedancias por las que la fuente induce tensiones y/o corrientes sobre la víctima.

Impedancia entre líneas

En la figura 28.9 se muestra un caso en el que dos líneas no están teóricamente conectadas entre sí, sino que las separa un aislante; en realidad, ningún aislante es perfecto y puede presentar una resistividad más o menos alta, pero no infinita. La resistencia R modela ese efecto.

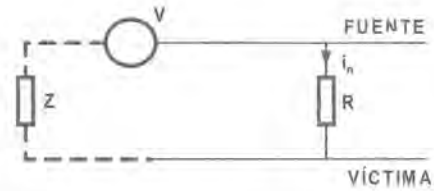


Figura 28.9. Acoplamiento por resistencia entre líneas.

La corriente se cerraría sobre una impedancia Z que conectaría por otro punto la fuente y la víctima. Aunque esto pueda parecer algo poco frecuente, lo cierto es que se produce muchas veces, sobre todo si se tiene una fuente sometida a tensiones muy elevadas que haría que la corriente de ruido i_n pudiese ser significativa a pesar de altos valores de la resistencia entre las líneas. La impedancia Z de conexión suele introducirse en muchos casos a través de los transformadores de las fuentes de alimentación cuyas capacidades parásitas conectan el primario con el secundario. En la figura 28.10 se muestra un ejemplo de acoplamiento por resistencia entre líneas entre dos circuitos teóricamente aislados.

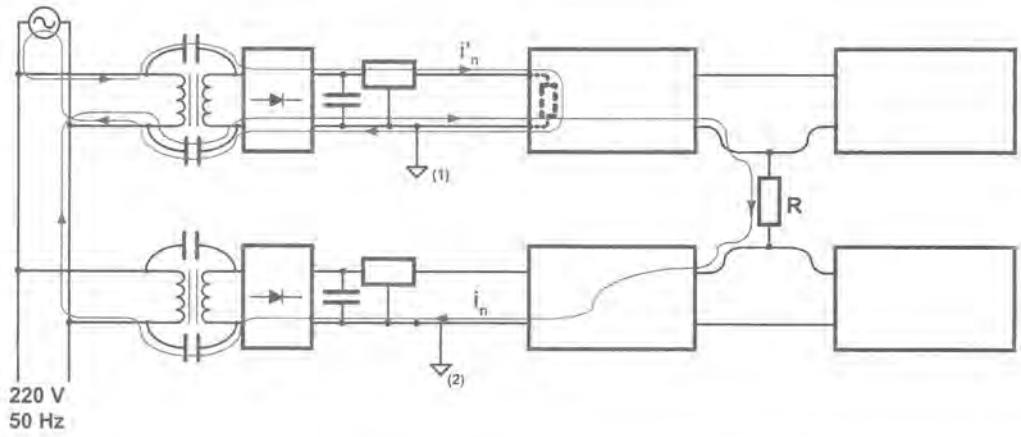


Figura 28.10. Acoplamiento por resistencia entre líneas entre dos circuitos teóricamente aislados.

Aunque en el circuito de la figura 28.10 se ha dibujado uno de los posibles caminos de cierre de la corriente de ruido i_n, otros más son posibles, alguno de ellos introduciendo, además, la tensión de la red eléctrica en el camino de circulación. En este caso, parte del acoplamiento no es resistivo ya que el cierre del circuito se produce por medio de las capacidades parásitas de los transformadores de las fuentes de alimentación; este efecto podría producirse en uno de los circuitos ya que sería factible que se cerrase corriente por las capacidades parásitas del transformador a través de cualquier componente situado entre el

positivo y la masa de alimentación (todo el circuito) tal y como se indica en el circuito dibujado a puntos que conduce la corriente de ruido i_n'.

En general, los valores de estas corrientes suelen ser muy débiles en relación con otros tipos de acoplamiento y rara vez se tiene en cuenta excepto que la tensión entre los dos sistemas sea muy alta como es el caso de circuitos electrónicos que estén próximos a redes de alta tensión.

Impedancia común

Esta es la forma más frecuente de acoplamiento conductivo y se produce cuando dos circuitos comparten una impedancia Z por la cual circulan corrientes procedentes de cualquiera de ellos, tal y como se indica en la figura 28.11.

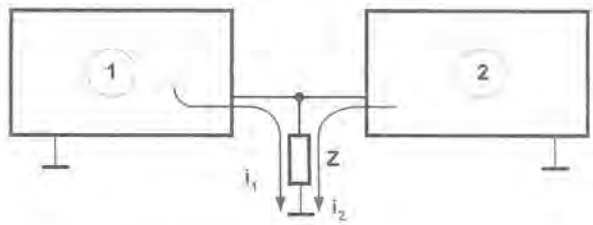


Figura 28.11. Interferencia mutua entre dos circuitos producida por impedancia común.

La tensión producida sobre la impedancia Z se debe a la suma de la producida por el circuito 1 y la producida por el circuito 2:

$$v_Z = Z(i_1 + i_2)$$

Por lo tanto, cada uno de los circuitos "ve" una tensión sobre la impedancia Z diferente de la esperada y, en función de los valores de las corrientes, uno de ellos será más o menos perturbado que el otro. Aunque el efecto es mutuo, se considerará como fuente de interferencia aquel cuyo valor de corriente sea mayor.

Para que este tipo de acoplamiento se produzca, los dos circuitos deben compartir algún punto (en la figura 28.11 se suponía que era la masa por simplificar pero puede ser cualquiera) y es un efecto frecuente en sistemas electrónicos que tengan bloques diferenciados (analógicos, digitales y de potencia) pero conectados entre sí; en estos casos, hay que cuidar la aparición de este fenómeno para que las etapas de potencia no perturben a los bloques analógicos y digitales y para que estos últimos no introduzcan señales no deseadas en las etapas analógicas.

Teniendo en cuenta que en la mayoria de los sistemas electrónicos suelen existir diversos tipos de bloques conectados entre sí, el diseño del conexionado del conjunto debe cuidarse para evitar la presencia de estos problemas. En los dos capítulos siguientes se tratará de reducir este efecto mediante un correcto diseño del cableado y de los circuitos impresos.

Otra forma de interferencia por impedancia común es la que se produce en las conexiones de alimentación de los diversos bloques de un sistema. En la figura 28.12 se muestra un caso en el que dos bloques comparten la misma alimentación.

La tensión de alimentación que "ven" ambos circuitos sería la obtenida de:

$$V'_{CC} = V_{CC} - (2R_L + R_S)(i_1 + i_2) - 2L_L\left(\frac{di_1}{dt} + \frac{di_2}{dt}\right) \qquad (28.1)$$

donde R_L y L_L son la resistencia e inductancia de los cables de alimentación y R_S es la resistencia de salida de la fuente de alimentación; como se puede observar de (28.1), ambos circuitos perturban los valores de la tensión de alimentación y, lo que es peor, modifican el valor de la tensión de referencia de su masa (punto G en la figura 28.12) que ya no será cero voltios.

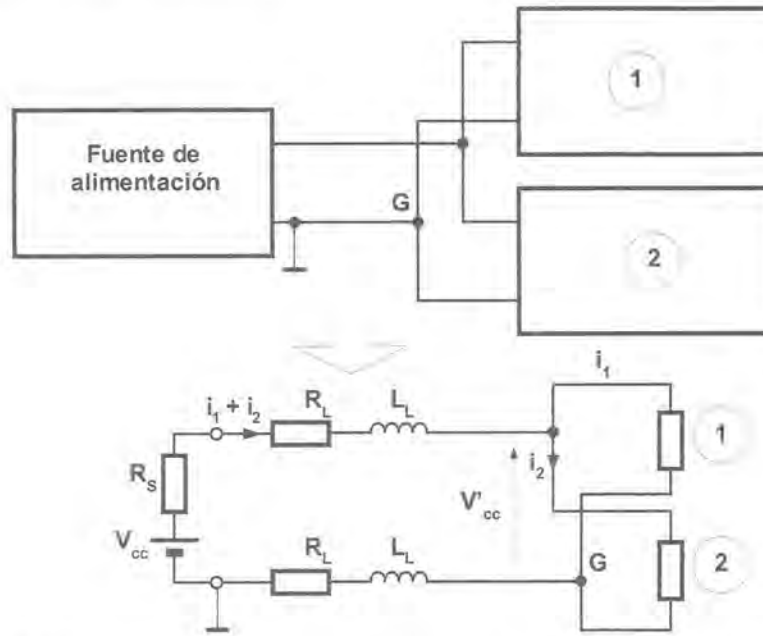


Figura 28.12. Problemas de impedancia común en un típico ejemplo de cableado de alimentación.

El problema se complica cuando los consumos son pulsantes lo que origina fuertes caídas de tensión en las líneas de alimentación. En este caso, se deben situar condensadores de desacoplo en la entrada de cada bloque que sean capaces de absorber los picos de corriente y evitar que estos se propaguen y generen dificultades a otros circuitos o al propio bloque[4].

Cuando uno de los circuitos presenta un consumo muy superior al otro, el de consumo más elevado puede afectar al de consumo más bajo, provocando una caída de tensión en la alimentación que no sólo introduce ruido, también podría hacer que uno de ellos o ambos no funcionasen correctamente si la expresión (28.1) arrojara valores fuera del margen de tensión permitida.

Para reducir el efecto de las impedancias comunes en los cables de alimentación se pueden tomar algunas iniciativas:

a) Reducir la distancia entre la fuente de alimentación y los puntos de consumo con lo que caerán los valores de la resistencia y de la inductancia asociadas a la línea.

b) Incrementar la sección de los conductores lo que hará disminuir el valor de la resistencia del hilo.

c) Cablear la alimentación en estrella, con un par de conductores para cada uno de los bloques de consumo con lo que el camino común se reduce a sólo la resistencia de salida de la propia fuente que suele ser de muy bajo valor.

[4] En el caso de circuitos de tipo CMOS cuyo consumo es siempre pulsante (sólo absorben corriente en los momentos de conmutación) si no se sitúa el condensador de desacoplo, el circuito llega a no funcionar porque la tensión de alimentación cae por debajo de los valores permisibles en algunos momentos.

d) Disponer condensadores en todos los puntos de consumo lo que reduce los picos de corriente y hace que el valor de las corrientes de consumo sea mucho más constante con lo que se elimina de facto el efecto de las inductancias asociadas a las líneas de cableado.

Aunque algunas veces es fácil seguir estas recomendaciones, como en el caso del cableado de la alimentación de un PC (en estrella y con cables de alta sección) en otras ocasiones no resulta tan evidente; en los dos capítulos siguientes se comentarán los procedimientos para llevar a cabo de la mejor manera posible estas sugerencias en el caso del cableado de sistemas y en el diseño de tarjetas de circuito impreso.

28.3.2 ACOPLAMIENTOS NO CONDUCTIVOS

Supongamos dos conductores separados entre sí y entre los cuales no hay impedancias comunes ni caminos explícitos o implícitos que permitan la circulación de corriente y consideremos que uno de ellos —la fuente de interferencias— está sometido a unas condiciones eléctricas determinadas que se traducen en una tensión y una corriente. En estas circunstancias, la fuente de interferencias puede inducir sobre el otro conductor —*la víctima*— *tensiones y corrientes sin el concurso de impedancias comunes debidas a los siguientes efectos* (figura 28.13) [28.11]:

a) La fuente genera un campo eléctrico que afecta a la víctima. A menudo, este efecto se denomina interacción electrostática o acoplamiento electrostático ya que se debe a un campo eléctrico; sin embargo, el efecto se produce sea o no sea estático el campo producido por lo que aquí usaremos el término más correcto de *acoplamiento eléctrico o capacitivo.*

b) La fuente genera un campo magnético que afecta a la víctima. En algunos textos se nombra este efecto como interacción electromagnética o acoplamiento electromagnético aunque se produce sin el concurso de ningún campo eléctrico; por ello, aquí lo denominaremos *acoplamiento magnético o inductivo.*

c) La combinación de los efectos del campo magnético y eléctrico producidos por la fuente también puede afectar a la víctima en lo que se denomina —aquí sí que parece que todos están de acuerdo— *radiación electromagnética.*

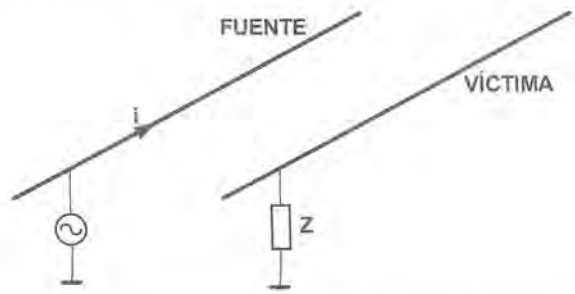


Figura 28.13. Acoplamiento no conductivo entre fuente y víctima.

En lo que sigue consideraremos que las corrientes y tensiones inducidas sobre la víctima no generarán nuevos campos ni afectarán al campo producido por la fuente ya que su valor resultará muy pequeño en relación con aquel. También consideraremos que las longitudes de los cables son mucho menores que la longitud de onda correspondiente a la máxima frecuencia de señal que circule por la fuente.

En estas condiciones podemos preguntarnos cómo considerar los efectos de cada uno de los mecanismos de acoplamiento.

La respuesta dependerá, en general, de la propia fuente, del medio que la rodea y del punto en el que queramos conocer el efecto:

a) En las proximidades de la fuente los efectos de los campos dependen fundamentalmente de las características de la fuente (*campo cercano*).

b) En distancias grandes, el efecto de los campos depende principalmente del medio por el que se propaga la señal (*campo lejano*).

La *impedancia de la onda*[5] Z es el cociente entre el valor del campo eléctrico E generado por la fuente y el valor del campo magnético H según la definición de MIL-STD 188, es decir,

$$Z = \frac{E}{H} \tag{28.2}$$

El valor de Z en función de la distancia a la fuente está dibujada en la figura 28.14. La curva superior representa el caso de un emisor de campo eléctrico fundamentalmente, generando poco campo magnético, mientras que la curva inferior presenta el contrario. En ambos casos, se está en consideración de campo cercano, mientras que cuando los dos campos adquieren consideración similar, se tiene la tercera parte de la curva que es una recta plana e independiente de la distancia a la fuente (campo lejano). En este caso, la expresión (28.2) se transforma en:

$$Z = \sqrt{\frac{\mu}{\varepsilon}} \tag{28.3}$$

donde μ es la permeabilidad magnética del medio y ε, su permitividad dieléctrica. Para el aire o el vacío (para el que se ha dibujado la curva en la figura 28.14) el valor de μ es de $4\pi 10^{-7}$ H/m y el valor de ε es de $10^{-9}/36\pi$ F/m con lo que $Z = 120\pi \approx 377\ \Omega$.

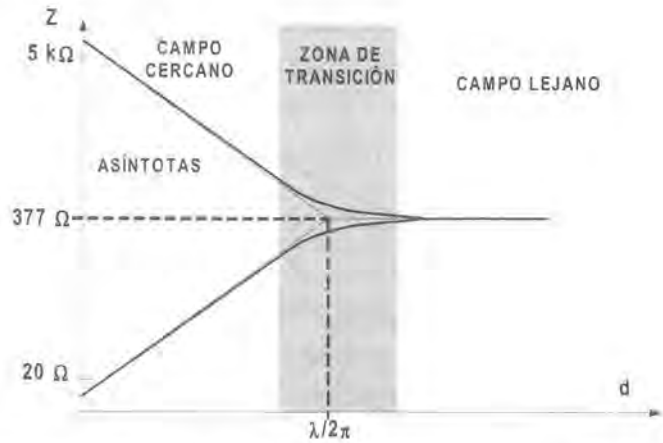


Figura 28.14. Impedancia de onda en función de la distancia que define las zonas en que se puede considerar el campo cercano o lejano.

El punto en que cambia la consideración de campo cercano a campo lejano no es un lugar exacto, sino que es una zona de transición tal y como se muestra en la figura 28.14; sin embargo, considerando el punto en que se cruzan las dos asíntotas de las curvas de predominio del campo eléctrico y magnético, resulta a una

[5] La impedancia de la onda es, en realidad, la impedancia del espacio que rodea a la fuente.

distancia de $\lambda/2\pi$ siendo λ la longitud de onda de la señal; como quiera que la longitud de onda depende la frecuencia f de la señal según $\lambda = c/f$ donde c es la velocidad de la luz, resulta que la distancia crítica d_c es:

$$d_c = \frac{c}{2f\pi} \tag{28.4}$$

y, para el vacío (o el aire), $c \approx 300000$ km/s, resulta que:

$$d_c = \frac{47{,}8 \cdot 10^6}{f} \tag{28.5}$$

Esto significa que, para señales de una frecuencia de 100 MHz que podrían circular por un sistema digital cualquiera, por ejemplo, la distancia crítica sería de algo menos de medio metro y la consideración de campo cercano sólo tiene validez en las proximidades más inmediatas del sistema. Sin embargo, cuando la frecuencia es menor, como en la mayoría de los sistemas electrónicos que no sean de comunicación o digitales de alta frecuencia, las distancias para considerar el campo cercano son mucho mayores. Un sistema que contenga frecuencias de 1 MHz, por ejemplo, presentaría la zona de transición entre campo cercano y lejano en el entorno de 50 metros.

Seguidamente vamos a considerar el efecto del campo cercano que se debe a los dos mecanismos de acoplamiento: el capacitivo y el inductivo.

Acoplamiento capacitivo

El acoplamiento capacitivo que tiene lugar entre dos conductores puede presentarse como se muestra en la figura 28.15 que corresponde a la más sencilla de sus expresiones.

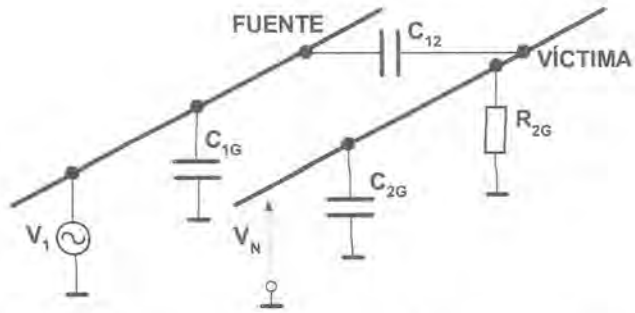


Figura 28.15. Capacidades relevantes en el acoplamiento capacitivo.

En esta figura se muestran dos conductores, uno de los cuales actúa como fuente de interferencias y el otro como víctima de las mismas. Las capacidades que se indican en la figura corresponden a los parásitos capacitivos entre el conductor fuente y masa (C_{1G}), entre la víctima y masa (C_{2G}) y la que corresponde a la presente entre ambos, C_{12}, mientras R_{2G} es la resistencia que haya conectada entre la víctima y masa y resultará el paralelo de las resistencias de todos los componentes conectados entre el conductor víctima y masa.

La fuente V_1 provocará en el conductor víctima una tensión de ruido, V_N que vendrá determinada por el divisor de impedancias formado y que resultará ser:

$$V_N = \frac{j\omega R_{2G} C_{12}}{1 + j\omega R_{2G}(C_{12} + C_{2G})} V_1 \tag{28.6}$$

La expresión (28.6) se simplifica cuando se cumpla que la frecuencia de la señal sea baja, es decir,

$$1 >> j\omega R_{2G}(C_{12} + C_{2G})$$

en cuyo caso, la tensión de ruido inducida sobre el conductor víctima sería:

$$V_N = j\omega R_{2G} C_{12} V_1 \qquad (28.7)$$

o, lo que es lo mismo, el efecto de una hipotética fuente de corriente de valor $j\omega C_{12}V_1$ sobre la resistencia que hay entre el conductor víctima y masa, R_{2G}. El valor de C_{12} dependerá de la distancia que separe los conductores, disminuyendo en la medida que esta crece. En la gráfica de la figura 28.16 se muestra el valor de la tensión de ruido en función de la frecuencia de la señal de la fuente.

Para frecuencias bajas —sirve la expresión (28.7)— la tensión de ruido se minimizaría al reducir el valor de la resistencia que hay entre el conductor víctima y masa pero, a frecuencias altas, la fuente de tensión resulta independiente de la carga y sólo interviene el divisor capacitivo por lo que una reducción en el valor de R_{2G} no tendría demasiado efecto sobre el nivel de ruido inducido.

La forma de reducir el efecto del acoplamiento capacitivo es la utilización de una pantalla eléctrica que rodee al conductor víctima: en la figura 28.17 se muestra un cable apantallado y el correspondiente equivalente eléctrico.

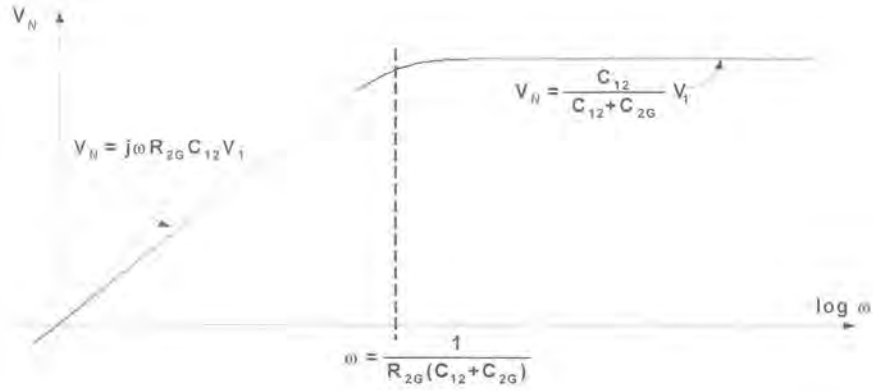


Figura 28.16. La tensión de ruido crece con la frecuencia de la fuente de interferencia hasta que, sobrepasado un determinado valor de frecuencia, permanece casi constante.

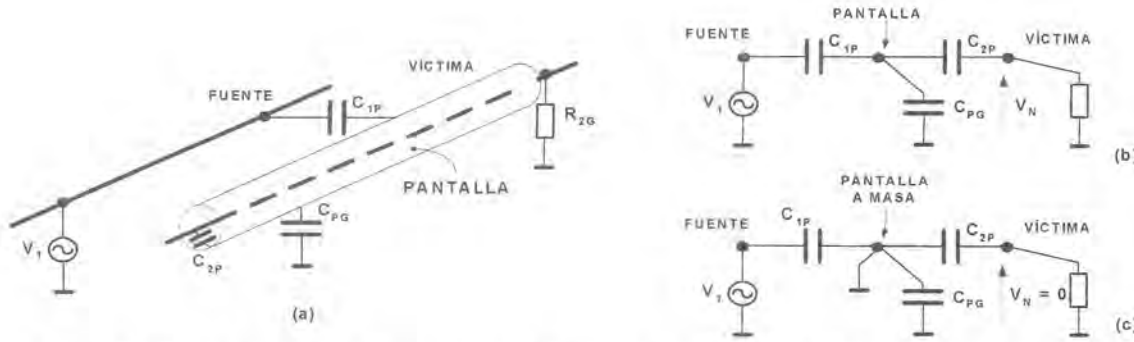


Figura 28.17. Empleo de una pantalla para la reducción del efecto del acoplamiento capacitivo.

En el circuito equivalente de la figura 28.17 podría calcularse el valor de la tensión de ruido en la víctima obteniendo el valor de V_N. Sin embargo, la utilidad de la pantalla es la conexión a masa de la misma con lo que se garantiza que la tensión de ruido en ella es nula y, con ello, la tensión inducida sobre la víctima (figura 28.17c). En la práctica, la pantalla no cubre totalmente el cable, sino que deja al descubierto las conexiones de los extremos por lo que no se consigue un apantallado completo; por otra parte, de lo buena que sea la conexión a masa de la pantalla depende la tensión que soporte esta y, con ello, la tensión de ruido que verá el conductor víctima. Así pues, habrá que evitar que la tensión de la pantalla fluctúe porque pasaría a ocupar el papel de la fuente de interferencia, llevando sus fluctuaciones a la víctima a la que trata de proteger según se muestra en la figura 28.18.

En este caso, la tensión de ruido que se desarrollaría sobre el conductor víctima sería:

$$V_N = \frac{j\omega C_{2P} Z}{j\omega C_{2P} Z + 1} V_P \qquad (28.8)$$

expresión que resulta muy parecida a la (28.6) y sobre la que cabe hacer las mismas consideraciones; sin embargo, hay que hacer notar que el valor de C_{2P} puede ser conocido ya que un conductor apantallado presenta un equivalente capacitivo que se asemeja bastante al de un condensador cilíndrico.

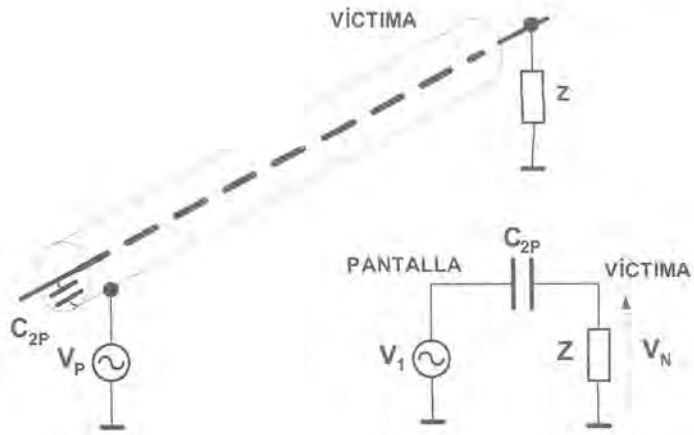


Figura 28.18. Si el apantallamiento no es correcto porque la tensión en la pantalla no sea constante, se introducen señales en el cable que intenta proteger.

La eficacia de la protección de la pantalla depende del buen conexionado de los cables; a este aspecto se dedicará el siguiente capítulo, mientras que la consecución de estos objetivos en las tarjetas de circuito impreso será objeto del capítulo 30.

Acoplamiento inductivo

La presencia de un conductor por el que circula una corriente genera a su alrededor un campo magnético *cuyas variaciones pueden ocasionar circulaciones de corriente por la malla de un circuito víctima según se* muestra en la figura 28.19.

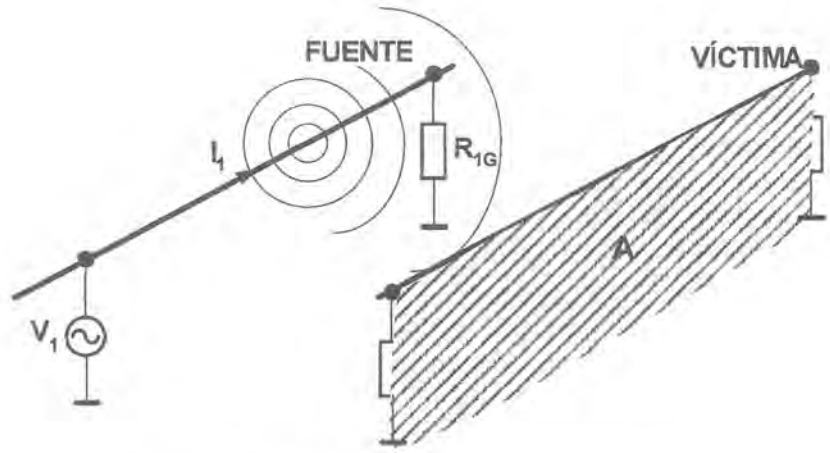


Figura 28.19. Causa del acoplamiento inductivo.

Sobre el circuito víctima se inducirá tensión siempre que haya una variación del flujo cortado por el área A, lo que en la práctica significa:

a) Que el campo generado por el conductor fuente sea variable lo que sería consecuencia de una corriente variable circulando por ese conductor.

b) Que alguno de los dos conductores —fuente o víctima— se mueva modificando el número de líneas de flujo que corta el área A.

c) Las dos circunstancias anteriores de forma más o menos simultánea.

En cualquier caso, siempre que haya variación de flujo se inducirá una tensión de ruido, V_N en la malla de área A según la expresión:

$$V_N = -\frac{d\left(\int_A \vec{B}\cdot\overrightarrow{dA}\right)}{dt} \tag{28.9}$$

Si la malla de área A y la fuente de interferencia conservan su posición relativa y la corriente I_1 varía senoidalmente, el campo producido también varía de la misma forma y la tensión inducida resulta ser:

$$V_N = j\omega BA\cos\theta \tag{28.10}$$

siendo θ el ángulo que forma la normal a la malla con el campo B. El modelo de interferencia corresponde esencialmente con el de un acoplamiento entre dos inductancias con un coeficiente de acoplamiento mutuo M_{12}[6] (figura 28.20). Teniendo en cuenta este coeficiente, el valor de la tensión de ruido vendrá dado por:

$$V_N = M_{12}\frac{dI_1}{dt} \tag{28.11}$$

[6] El cálculo de la inductancia mutua M_{12} puede llegar a ser muy complejo ya que dependerá de la geometría de la fuente y del circuito víctima. En [28.11] se indica la forma de calcularla para casos sencillos.

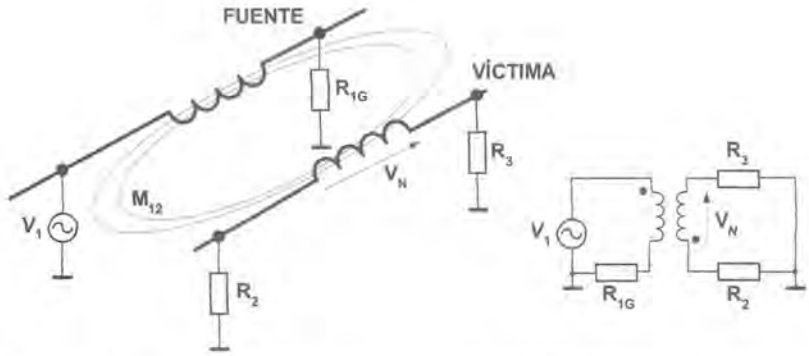


Figura 28.20. Acoplamiento inductivo y circuito eléctrico equivalente.

De la observación de la figura y de las expresiones (28.9) y (28.10) se puede concluir que el modelo de ruido introducido por el acoplamiento inductivo corresponde al de una fuente de tensión proporcional al campo producido.

La reducción del ruido presente en la víctima por acoplamiento inductivo se puede llevar a cabo incidiendo sobre los parámetros que afectan a la tensión, según las expresiones (28.9) y (28.10), es decir:

a) La distancia entre fuente y víctima es un factor crucial ya que el campo B se amortigua con ella; por lo tanto, separar los dos circuitos es una práctica que reduce el acoplamiento.

b) La disminución del área de la malla del circuito víctima reducirá notablemente el ruido inducido; esto se consigue mediante el trenzado de los cables y su adecuada conexión según se comentará en *el siguiente capítulo o con diseños correctos de las tarjetas de circuito impreso (véase capítulo 30).*

c) La reorientación de la superficie de la malla para conseguir que los vectores de campo B y superficie A sean perpendiculares (o lo más perpendiculares que sea posible). Esto es difícil de conseguir en la mayoría de los casos puesto que en diseños generales no es fácil conocer exactamente la distribución de las líneas de campo magnético. Por ejemplo, en un equipo los cables de alimentación deben tenderse en perpendicular a las pistas y cables de señal.

d) Evitar cualquier movimiento relativo (como vibraciones, por ejemplo) ya que la variación del producto escalar BdA resultaria distinta de cero incluso con campos fijos como los generados por corrientes continuas.

En general, una fuente de interferencias generará problemas por los dos mecanismos estudiados aquí, es decir, mediante acoplamiento inductivo y mediante acoplamiento capacitivo por lo que el ruido introducido externamente a un circuito se podrá considerar siempre como la suma de los efectos ocasionados por una fuente de tensión de ruido y una fuente de corriente de ruido tal y como se muestra en el modelo de la figura 28.21, justificando el modelo que se empleó en el capítulo 2 (figura 2.3).

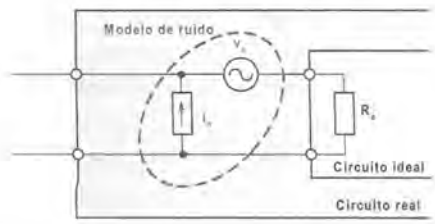


Figura 28.21. Modelo final de ruido, modelado como una fuente de tensión en serie *con la entrada y una fuente de corriente en paralelo.*

Radiación electromagnética

La generación de interferencias ante la presencia de campos electromagnéticos resulta de una combinación de los efectos de la componente eléctrica y la de la componente magnética por lo que el tratamiento resulta similar al estudiado hasta aquí, en el fondo no se trata de ninguna nueva propiedad sino una combinación de efectos conocidos.

Por tanto, la radiación electromagnética no presenta mayores particularidades para tenerla en cuenta en lo que a la generación de ruido se trata. La única salvedad radica en el valor de los campos presentes cuya variación con la distancia puede presentar algunas particularidades: en efecto, cuando las frecuencias son elevadas, los cables de longitud apreciable en relación con la longitud de onda de las señales, las consideraciones hechas hasta este momento pierden validez y pueden aparecer fenómenos diferentes como cancelaciones del campo en una zona concreta [28.11]. Sin embargo, todos estos factores afectan al valor del campo en un punto concreto y no al efecto que ese campo tendrá sobre el circuito víctima que, cualitativamente hablando, será similar al efecto de un campo magnético y uno eléctrico.

Aunque la naturaleza del fenómeno pueda establecer algunas particularidades cuantitativas, básicamente los mecanismos para paliar los efectos de la radiación electromagnética son los mismos que ya se han detallado antes y la única precaución que habrá que añadir es la emanada de los circuitos que pueden actuar como fuentes de interferencias de este tipo y que son:

a) Transmisores de radiofrecuencia.
b) Generadores de microondas.
c) Circuitos digitales de alta frecuencia.

Aunque la problemática de todos estos circuitos pueda parecer similar, los dos primeros se caracterizan por presentar espectros de emisión relativamente estrechos por lo que las interferencias generadas resultan más previsibles. En cambio, aunque los circuitos digitales de alta frecuencia suelen utilizar frecuencias fijas de reloj, la forma de onda de sus señales genera espectros de frecuencia más complejos y, además no tan previsibles por lo que se convierten en sistemas muy molestos desde el punto de vista de generación de interferencias. Las frecuencias de trabajo de estos circuitos, cada vez mayores, hace que los "cables de conexión" que constituyen las pistas del circuito impreso actúen como auténticas antenas radiando al ambiente una cantidad de energía respetable. El diseño de los sistemas digitales para prevenir este efecto es tan complejo como el propio diseño del circuito [28.7] y escapa bastante del ámbito de este libro.

28.4 Minimización de los efectos de las interferencias

Las interferencias de cualquiera de los tipos estudiados hasta aquí dan lugar a la aparición de tensiones y/o corrientes de ruido en los sistemas, produciendo una inevitable disminución de la relación S/N del sistema. Lógicamente, ese efecto será mayor en la medida que las señales de ruido inducidas sean grandes o que las señales del sistema víctima sean débiles. Por lo tanto, dada una hipotética fuente de interferencias presente en las proximidades de nuestro circuito, la pregunta que nos deberíamos formular es: ¿cómo conseguir que nuestro circuito sea lo más insensible que sea posible a las interferencias?

La respuesta a esta pregunta no resulta obvia, ni mucho menos, sino que entraña una dificultad elevada ya que las interferencias se deben a campos que resultan difíciles de modelar en cada caso y que dependen de factores tales como la proximidad de objetos metálicos o magnéticos, la geometría del circuito fuente, del circuito víctima...

En este caso, la validez de las ecuaciones presentadas aquí resulta bastante escasa a nivel cuantitativo y sólo deben ser tomadas como meras indicaciones de tendencias: si acercamos o alejamos esto, ¿qué ocurre?,

si aumentamos esta corriente o bajamos esta impedancia, ¿mejorará o empeorará el comportamiento del circuito respecto a las interferencias? Preguntas de este tipo pueden contestarse parcialmente pero es muy complejo —y tiene poco sentido— realizar cálculos muy exactos del nivel de ruido que se va a inducir.

Sin embargo, tratar de minimizar los efectos de las interferencias mediante la modificación de los circuitos potencialmente víctimas —todos— es una tarea muy importante ya que permite mejorar su compatibilidad electromagnética. Para ello, hay que identificar los puntos más problemáticos y tomar las oportunas medidas para reducir el efecto de las interferencias: desde este punto de vista, tenemos:

a) *Bloques que manejen señales débiles y que las señales inducidas se traducirán en una sensible* disminución de la relación S/N.

b) Cables de conexión entre circuitos, sistemas o bloques que, por ser conductores de longitudes más grandes se verán más afectados.

En algunos casos, hay etapas que reúnen ambas particularidades, incluyen cables y conexiones más o menos largos y manejan señales débiles con lo que resultan el caldo de cultivo más eficaz para una buena cosecha de ruido: estamos hablando de las etapas de entrada de los sistemas de instrumentación, aquellas que recogen señales débiles procedentes de los sensores y las transforman en señales que nuestro sistema va a procesar. Por ello, el cuidado de estas etapas va a ser crítico en el funcionamiento de los sistemas de instrumentación y de ellos va a depender, en gran medida, las prestaciones finales del equipo.

¿Cómo mejorar la compatibilidad electromagnética de estas etapas? A esta pregunta se dedicarán los capítulos siguientes en los que se tratarán los temas relacionados con el cableado y el diseño de las tarjetas de circuito impreso.

A grandes rasgos, consideraremos que nuestro circuito está en un ambiente hostil desde un punto de vista electromagnético y queremos hacer que sea lo más insensible a él que sea posible. Para ello, podemos recurrir a diversas técnicas:

a) Evitar que las señales de ruido se cuelen en nuestro sistema, lo que se consigue de tres formas:
   1. Con un correcto cableado e interconexión entre los componentes del sistema de forma que sean menos susceptibles a las señales exteriores.
   2. Mediante un adecuado diseño de las alimentaciones del circuito evitando las impedancias comunes entre bloques.
   3. Con la disposición de pantallas que modifiquen la distribución del campo eléctrico o magnético y preserven nuestro circuito o nuestras señales de esos campos.

b) Como quiera que va a resultar imposible conseguir eliminar todas las señales de ruido externo, habrá que reducir su efecto sobre el funcionamiento del circuito para lo cual podemos recurrir a:
   1. Reducir el ancho de banda al estrictamente necesario para el procesado de la información, evitando anchos de banda superiores ya que sólo constituyen caminos fáciles para las señales no deseadas.
   2. Emplear dispositivos capaces de absorber determinadas perturbaciones (condensadores, supresores de transitorios, *chokes*, etc.).
   3. Diseñar con robustez el sistema de modo que el funcionamiento pueda ser factible incluso en condiciones de ruido elevado.

Por otro lado, cualquier circuito es una potencial fuente de interferencias para los demás por lo que una forma de reducir los problemas mencionados hasta aquí consiste en evitar que se generen interferencias; dado que cada circuito trabajará como deba hacerlo, no suele ser posible modificar el modo de funciona-

miento de un equipo para hacer que produzca menos interferencias[7]. Sin embargo, lo que sí se debe hacer es procurar que la emisión de interferencias se reduzca mediante técnicas básicamente iguales a las empleadas para evitar que los campos eléctricos y magnéticos afecten a un sistema.

Resumen

Las interferencias de unos sistemas en otros son causas frecuentes de problemas en muchos sistemas electrónicos pero adquieren particular importancia en el caso de los sistemas de instrumentación electrónica por la presencia de señales débiles muy susceptibles de ser "molestadas" por fuentes externas. Así, una gran cantidad de los sistemas de instrumentación cabe calificarlos como víctimas de interferencias, sobre todo en sus etapas iniciales.

En este capítulo se han analizado las fuentes más comunes de perturbación, se ha indicado la normativa —muy abundante— sobre los distintos aspectos relacionados con la producción, propagación y efectos de las interferencias y, sobre todo, se ha introducido el concepto básico de compatibilidad electromagnética como una medida de la capacidad de un sistema de tolerar el funcionamiento en un ambiente determinado de interferencia electromagnética.

Uno de los aspectos más relevantes a la hora de estudiar el mecanismo de producción de interferencias es el acoplamiento de la energía desde el punto donde se emite (fuente) hasta el punto donde se recibe (víctima). El acoplamiento puede ser de tipo conductivo (por impedancia entre líneas o por impedancia común) y no-conductivo (capacitivo, inductivo o radiado). Estos acoplamientos dan lugar a la introducción de ruido en los sistemas que se puede, finalmente, modelar como una fuente de tensión y una fuente de corriente tal y como se había utilizado en capítulos anteriores.

Para concluir el capítulo se dan algunas ideas sobre la forma de minimizar las interferencias o disminuir sus efectos que pasa por las acciones de eliminación de las fuentes de ruido, reducción de los mecanismos de acoplamiento y la insensibilización de las víctimas potenciales.

[7] En algunas topologías de equipos de electrónica de potencia sí que puede modificarse su modo de funcionamiento para reducir la aparición de flancos bruscos en las conmutaciones haciendo que el sistema sea menos proclive a producir interferencias y, además, se reduzca el espectro de emisión de las mismas.

29

Cableado y apantallado

29.1 Minimización de interferencias conductivas
29.2 Minimización de interferencias debidas a acoplamientos inductivos
29.3 Minimización de interferencias debidas a acoplamientos capacitivos
29.4 Minimización de interferencias radiadas. Pantallas

En el capítulo anterior se han tratado los efectos de las interferencias electromagnéticas, así como los mecanismos básicos por los que se producen. El presente capítulo está dedicado fundamentalmente a cómo se debe realizar la conexión entre los diferentes bloques que componen un sistema de instrumentación para minimizar las interferencias, es decir, al cableado.

Para conectar, por ejemplo, un sensor que proporciona una determinada señal (en tensión, en corriente, etc.) con un instrumento cualquiera resulta obvio que se necesitan, en general, dos conductores o más (pensemos, por ejemplo, en una RTD conectada a tres hilos). Aparentemente, resulta muy sencilla la conexión: tan sólo serán necesarios unos cables con la longitud suficiente para cubrir la distancia que separa ambos sistemas; sin embargo, el correcto conexionado de los sistemas no es, ni mucho menos, tan obvio como pueda parecer. Cuanto más largo sea un conductor, más probable que sufra interferencias y mayores serán, y los cables de conexión entre los diferentes dispositivos son, generalmente, los conductores más largos que existen en un sistema de instrumentación por lo que será necesario cuidar el cableado con el fin de minimizar las interferencias.

29.1 Minimización de interferencias conductivas

Las interferencias conductivas se deben a que el sistema que actúa como fuente de la interferencia y la víctima comparten alguna impedancia. Los problemas más importantes se centran en las masas, en las alimentaciones (aunque en menor medida que las masas) y en la tierra.

29.1.1 CONEXIONADO DE LAS MASAS Y ALIMENTACIONES

El correcto conexionado de la masa resulta crítico puesto que es la referencia para todas las tensiones en los circuitos. Existen dos formas de conexionado:

- Masa en serie (figura 29.1a): es un tipo de conexionado poco recomendable. Así, en el ejemplo de la figura 29.1a, la tarjeta 3 provoca interferencias en el circuito 1 a través de la impedancia Z_1 del cable. De todas formas, es un conexionado muy simple de realizar por lo que se utiliza en aplicaciones poco críticas.
- Masa en paralelo (figura 29.1b): resulta mucho más recomendable porque se minimizan las impedancias compartidas (en el ejemplo de la figura quedarían reducidas a las pistas de masa de la fuente de alimentación), aunque es más costoso puesto que la longitud del cableado aumenta.

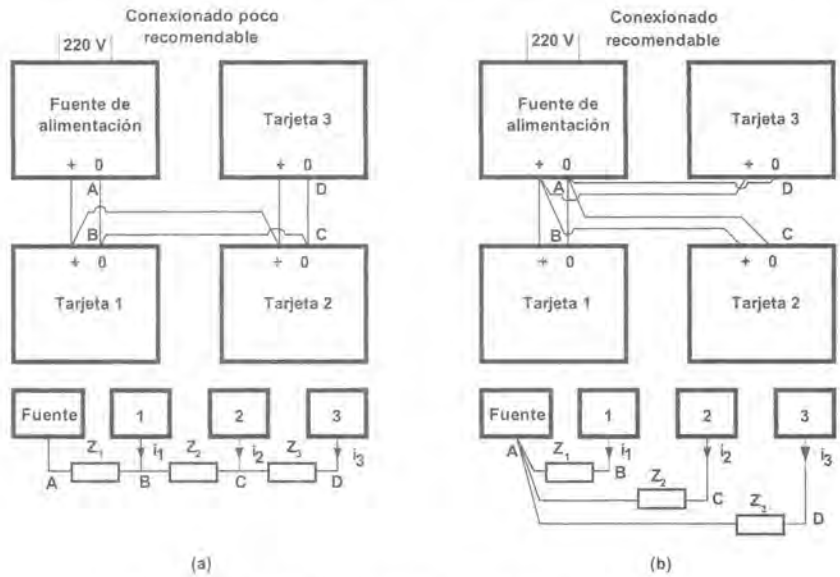


Figura 29.1. (a) Masa en serie; (b) masa en paralelo.

Las alimentaciones resultan menos críticas que las masas en cuanto que no son las referencias de las tensiones. De todas formas, las interferencias en las alimentaciones de las tarjetas provocan ruidos en los circuitos electrónicos que constituyen las tarjetas de circuito impreso. Conviene, por ello, realizar un cableado similar al de las masas en la medida de lo posible y utilizar condensadores de desacoplo a las entradas de las tarjetas (véase el siguiente capítulo).

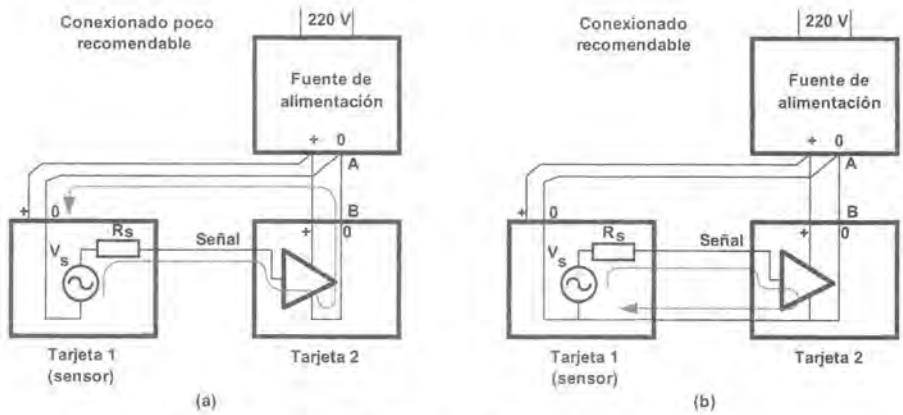


Figura 29.2. Conexión de señales entre tarjetas.

Por otra parte, cuando se transmiten señales entre equipos o tarjetas de circuito impreso, conviene conectar también la señal como la masa, a pesar de que esta última ya pueda estar conectada a través de las alimentaciones. Por ejemplo, en la figura 29.2 se pretende enviar la señal de un sensor incluido en la tarjeta 1 (se ha representado por su equivalente de Thévenin) a la tarjeta 2 cuya entrada es, por ejemplo, un amplificador. Si no se conecta la masa, además de la señal, el camino de retorno es muy largo y se ve interferido por las corrientes de alimentación de las tarjetas. Realizar la conexión de la masa permitirá, también, apantallar el cable y disminuir el acoplamiento capacitivo.

29.1.2 PROBLEMÁTICA DE LA CONEXIÓN A TIERRA

29.1.2.1 Bucles de tierra

Supóngase, por ejemplo, que se dispone de un sensor o un instrumento de medida cuyo equivalente es una señal V_s proporcional a una determinada magnitud que se pretende medir y una resistencia de salida R_s y que, además, por cualquier circunstancia, el circuito está puesto a tierra (figura 29.3a). Dicha señal se envía a través de unos cables más o menos largos a una sala de control hasta un equipo cuya entrada es, por ejemplo, un amplificador inversor como el de la figura 29.3a. El circuito también está puesto a tierra. Supóngase, que entre las tierras exista una diferencia de potencial V_T (se menciona en el capítulo 31 que la Tierra no es una superficie equipotencial; aunque es un buen conductor no es ideal y existe normalmente una diferencia de potencial impredecible entre dos puntos cualesquiera debida a corrientes de derivación, campos magnéticos, viento solar, rayos, etc.). Conviene aclarar que en el circuito de la figura 29.3a no parece posible que exista tal tensión entre las tierras de los circuitos 1 y 2 en cuanto que están unidos por el conductor A-D; sin embargo, el conductor presentará una cierta impedancia (olvídense los conductores ideales para comprender los bucles de tierra y las interferencias, en general).

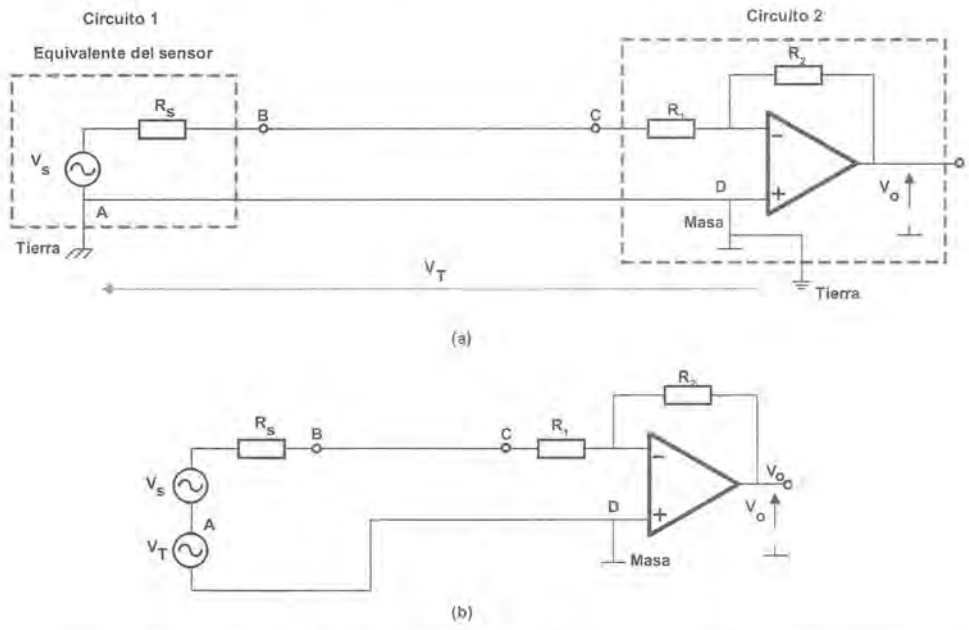


Figura 29.3. (a) Bucle de tierra; (b) circuito equivalente referido a la masa del circuito B.

Suponiendo que $R_s << R_1$, y que no existe diferencia de potencial entre las tierras, la tensión de salida V_o sería, idealmente:

$$V_o = -V_s \frac{R_2}{R_1} \tag{29.1}$$

Pero si se tiene en cuenta la diferencia de potencial V_T, el circuito equivalente referido a la masa del circuito 2 es el que se muestra en la figura 29.3b por lo que la tensión de salida es:

$$V_o = -(V_s + V_T)\frac{R_2}{R_1} \tag{29.2}$$

El error cometido dependerá del valor relativo de la diferencia de potencial entre las tierras, V_T, en relación con la tensión V_s, proporcionada por el sensor o el instrumento de medida. La realidad demuestra que la tensión V_T es impredecible, dependerá de la puesta a tierra, de las características del terreno, etc., pero podría estar comprendida entre algunos milivoltios hasta ¡¡¡decenas de voltios!!! si, por ejemplo, el circuito 1 y el 2 estuvieran situados en edificios diferentes. No sólo el error podría ser más que considerable, incluso se podría poner en riesgo el amplificador de entrada.

En la figura 29.4 se muestra otro equivalente eléctrico que, como se demostrará posteriormente, resultará más clarificador. La diferencia de potencial entre las tierras se traduce en una diferencia de potencial entre cada uno de los cables que conectan ambos circuitos y la masa del circuito 2. Es decir, mientras que la señal que se envía Vs, es una señal diferencial entre ambos conductores que da lugar a una corriente por el conductor BC i_d, que retorna por el conductor AD, la tensión en modo común V_T, da lugar a unas corrientes que circulan en el mismo sentido por ambos conductores, i_{cm1} e i_{cm2} y que retornan por la tierra. A este tipo de interferencias también se les denomina ruidos en modo común.

Cabe preguntarse por qué la señal diferencial retorna por el conductor y no por tierra. Existen dos motivos para ello: el primero es que el circuito que se ha representado en la figura 29.4a está un tanto simplificado; en realidad, existe una cierta impedancia en las conexiones a tierra y en la propia tierra por lo que el camino de retorno más sencillo es el conductor. Además, como los dos conductores estarán situados más próximos entre sí que con respecto a tierra, la inductancia mutua entre ambos conductores y la existente respecto a tierra es diferente y provoca que la corriente retorne básicamente por el conductor.

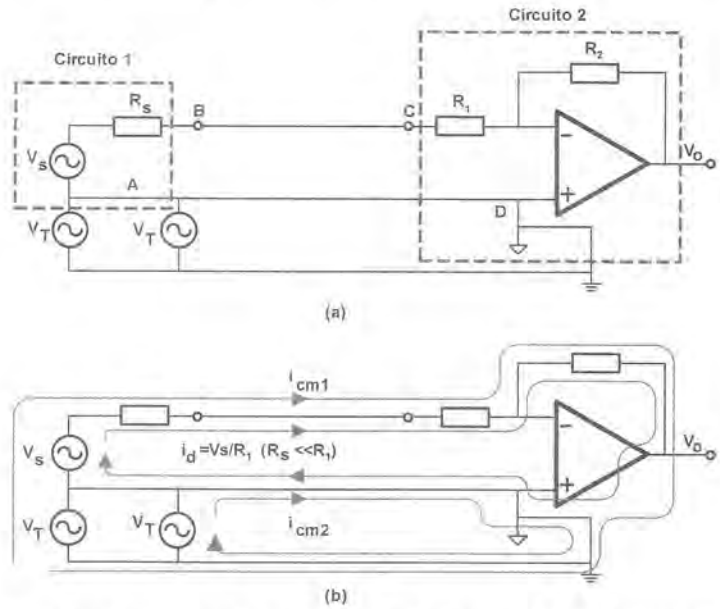


Figura 29.4. (a) Circuito equivalente al mostrado en la figura 29.3; (b) circulación de corrientes en modo común debido a los bucles de tierra.

Una solución muy sencilla al problema de las interferencias debidas a los bucles de tierra consistiría en realizar la conexión a tierra en un solo punto (figura 29.5); en este caso no existirían interferencias en modo común o serían muy pequeñas (téngase en cuenta que entre el punto A y la tierra existirá una determinada capacidad y, por tanto, un camino para la circulación de las corrientes en modo común, especialmente si la frecuencia es elevada).

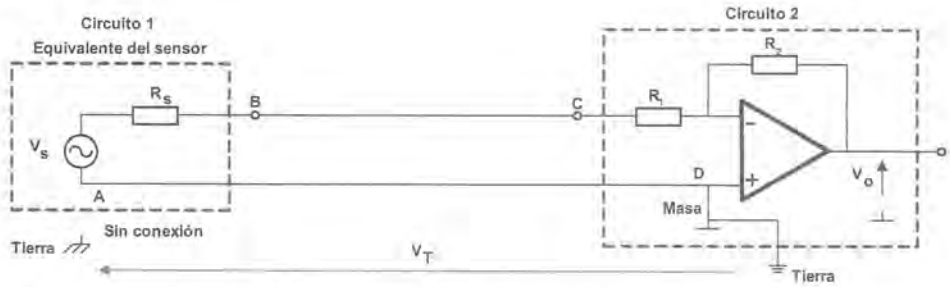


Figura 29.5. Solución al problema de los bucles de tierra realizando la conexión a tierra en un solo punto.

Sin embargo, en muchos casos esto no es posible por dos motivos:

- Porque por motivos de seguridad puede resultar obligatoria la conexión a tierra de ambos circuitos. Por ejemplo, si el circuito 1 está alimentado de la red eléctrica y tiene partes metálicas accesibles, las normas obligan a realizar la conexión a tierra. Si fuera simplemente un sensor, o un instrumento alimentado mediante baterías no sería necesaria tal conexión. Por los mismos motivos, el circuito de recepción estará en la mayor parte de los casos conectado a tierra.
- Porque se realiza una conexión a tierra inintencionadamente. Por ejemplo, si el circuito 1 es un termopar con la unión desnuda, cuando se pone en contacto con cualquier objeto apoyado sobre la tierra (como sucederá casi siempre), se está realizando una conexión a tierra. Lo mismo puede suceder cuando se emplean galgas extensiométricas, o sensores inmersos en depósitos de líquidos como los electroquímicos, de conductividad, etc.

EJEMPLO 29.1

Supóngase que se trata de medir el pH de una determinada solución mediante un sensor electroquímico correspondiente a un instrumento conectado a tierra. La solución se introduce en primer lugar en un recipiente metálico y se realiza la medición y en segundo lugar en un recipiente de vidrio que da lugar a un buen aislamiento con tierra. Si se observaran diferencias en la medición, podrían ser atribuibles a un bucle de tierra.

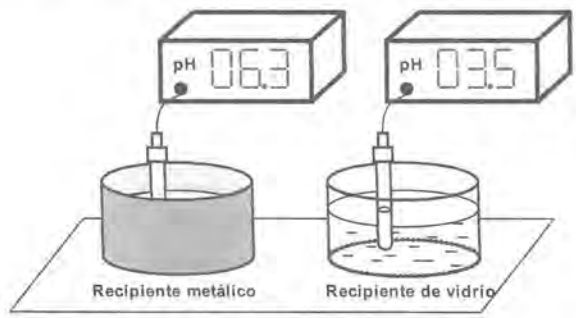


Figura 29.6

Por tanto, puede ser necesario recurrir a otras soluciones, entre las que se encuentran las siguientes:

- Emplear un amplificador diferencial.
- Aislar los circuitos.

Estas soluciones se describen a continuación.

(a) Entrada diferencial

Si el circuito de recepción fuera de tipo diferencial, como el amplificador de la figura 29.7a o un amplificador de instrumentación, el problema quedaría resuelto.

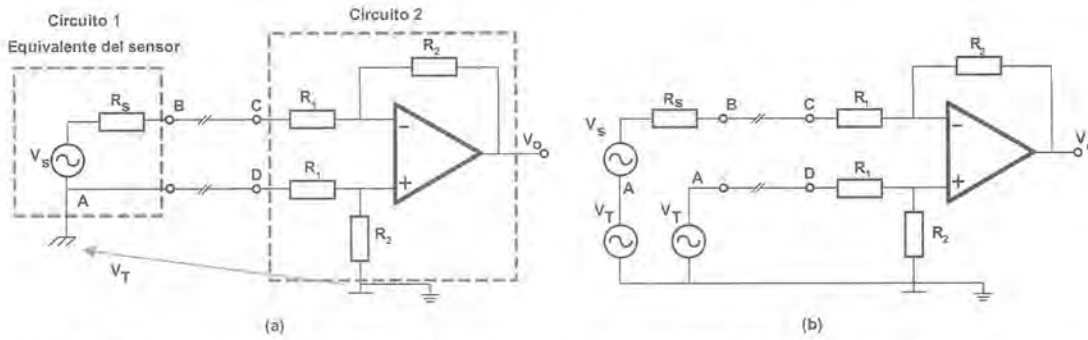


Figura 29.7. Solución al problema de los bucles de tierra empleando un amplificador diferencial.

En este caso la tensión de salida V_0 es igual a (figura 29.7b):

$$V_0 = -\left[(V_S + V_T) - V_T\right]\frac{R_2}{R_1} = -V_S\frac{R_2}{R_1} \tag{29.3}$$

por lo que sería independiente de la diferencia de potencial. Para ello, la razón de rechazo de modo común del operacional debería ser elevada. Los amplificadores operacionales presentan razones de rechazo de modo común muy grandes a bajas frecuencias pero, a partir de una determinada frecuencia, comienzan a disminuir notablemente. De todas formas, las interferencias debidas a los bucles de tierra son generalmente de baja frecuencia (típicamente de 50 Hz).

Conexión a tarjetas de adquisición de datos con entradas diferenciales

En una tarjeta de adquisición de datos existen, generalmente, entradas simples y entradas diferenciales, todas ellas multiplexadas. Por ejemplo, en la figura 29.8a se muestra el esquema básico de una tarjeta de cuatro entradas simples (CH1, CH2, CH3, CH4) y dos diferenciales (CH1 HI/CH1 LO, CH2 HI/CH2 LO). Para operar con la tarjeta es necesario configurar si el modo de funcionamiento es simple o es diferencial:

- Cuando la tarjeta se configura en modo simple, el amplificador de instrumentación tiene una de las entradas conectada a la masa de la tarjeta por lo que amplifica la tensión existente entre la conexión al canal que esté activado y masa. Por ejemplo, en la figura 29.8b se muestra la situación cuando se mide la señal del canal 1 en modo simple.
- Cuando se configura en modo diferencial, las entradas del amplificador de instrumentación están conectadas a dos de las líneas de entrada. Por ejemplo, en la figura 29.8c se muestra la situación cuando se mide en modo diferencial la señal del canal 1 (CH1 HI/CH1 LO).

A continuación, se describe cómo se debe realizar el correcto conexionado entre la señal y la tarjeta, y las ventajas e inconvenientes de cada modo [29.8].

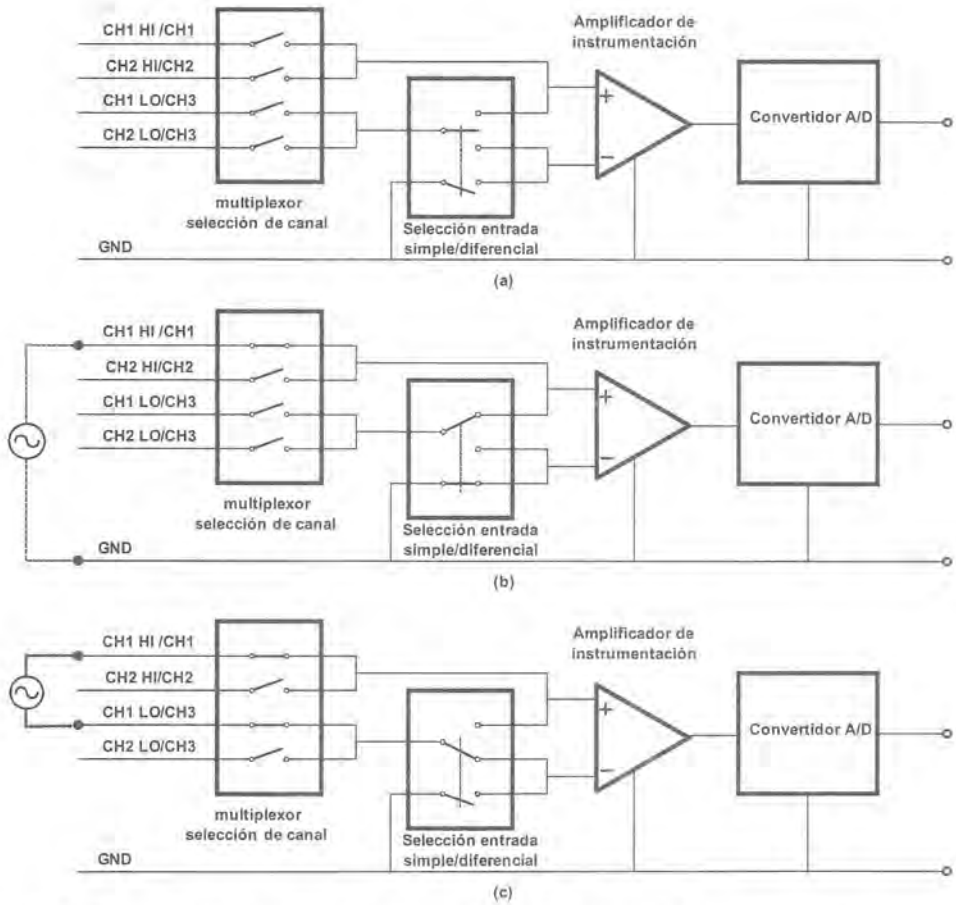


Figura 29.8. (a) Esquema básico de una tarjeta de adquisición de datos; (b) tarjeta configurada en modo simple para efectuar una lectura del canal 1; (c) tarjeta configurada en modo diferencial para efectuar una lectura del canal 1.

Modo simple

La conexión del circuito que proporciona la señal a la tarjeta de adquisición se realiza entre los terminales correspondientes al canal y a la masa de la tarjeta. Este modo de trabajo simple resulta adecuado cuando las señales que se pretende medir no están referenciadas a tierra (figura 29.9a), de lo contrario aparecen bucles de tierra (figura 29.9b) porque la tarjeta de adquisición también está conectada a tierra normalmente.

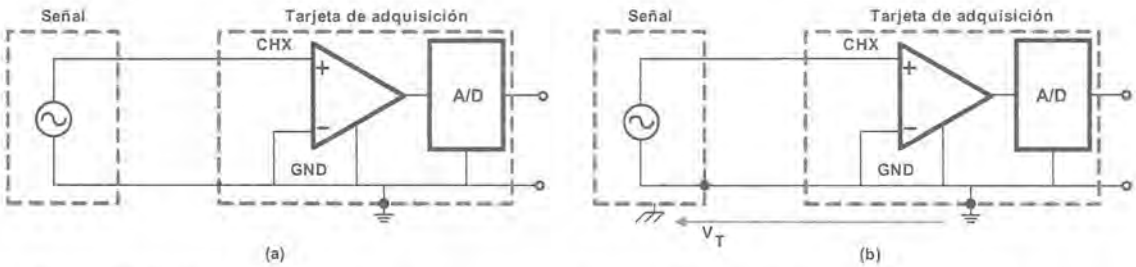


Figura 29.9. Conexión en modo simple. (a) Circuito de señal no referenciado a tierra; (b) circuito de señal referenciado a tierra (no recomendado debido a la aparición de bucles de tierra).

Modo diferencial

La ventaja de trabajar en modo diferencial (figura 29.10), como ya puede prever el lector, reside en que se pueden medir señales referenciadas a tierra sin que las interferencias en modo común den lugar a problemas debido a que la entrada es diferencial. El inconveniente es que el número de canales efectivos de la tarjeta queda reducido a la mitad.

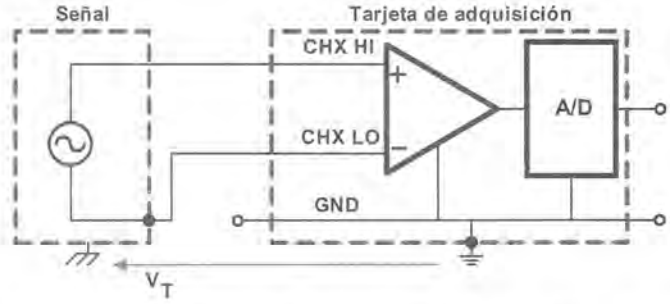


Figura 29.10. Conexión en modo diferencial cuando la fuente de señal está referenciada a tierra.

No obstante, cuando la señal no está referenciada a tierra es necesario tener ciertas precauciones debido a que la señal queda a un potencial indefinido respecto a la masa/tierra de la tarjeta de adquisición por lo que podría sobrepasarse la tensión de modo común que admite el amplificador de la tarjeta y dañarlo. Además, el amplificador de instrumentación queda sin camino para la circulación de las corrientes de polarización (figura 29.11a) por lo que mostraría un funcionamiento errático (probablemente se saturaría).

Para solucionar este problema resulta necesario añadir unas resistencias R_b entre las entradas de la tarjeta de adquisición y la masa (figura 29.11b). Sus valores óhmicos deberán ser lo bastante grandes como para no cargar excesivamente la fuente de señal y suficientemente pequeños como para no dar lugar a errores excesivos debidos a las corrientes de polarización. Suelen resultar adecuados valores comprendidos entre 10 kΩ y 100 kΩ.

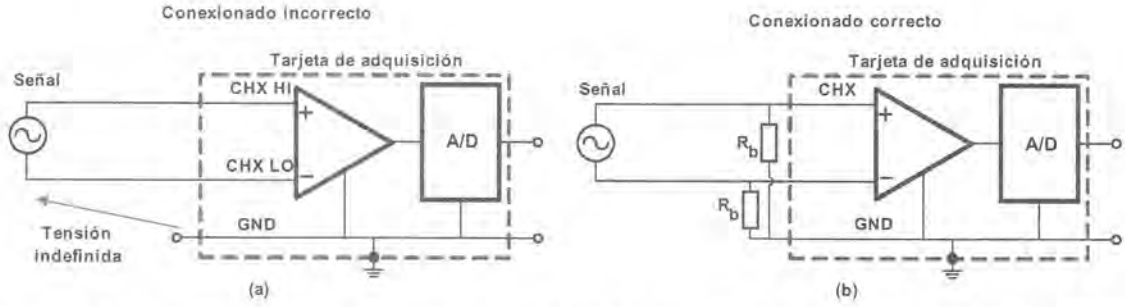


Figura 29.11. (a) Conexionado incorrecto de una señal flotante a una tarjeta de adquisición de datos funcionando en modo diferencial debido a que no existe camino para la circulación de las corrientes de polarización del amplificador; (b) conexionado correcto añadiendo unas resistencias de polarización.

Otra posibilidad consiste en conectar la masa del circuito de señal a la masa de la tarjeta de adquisición. Por ejemplo, en la figura 29.12 se muestra un posible conexionado de un circuito en puente de este modo. En este caso, el camino para las corrientes de polarización del amplificador de instrumentación lo provee inherentemente el circuito de señal por lo que no resultan necesarias resistencias adicionales.

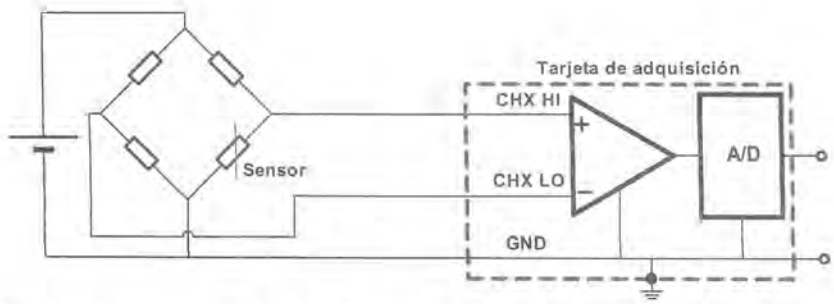


Figura 29.12. Conexionado en modo diferencial conectando, además, la masa del circuito correspondiente a la señal y la masa de la tarjeta de adquisición de datos.

(b) Aislamiento de los circuitos

La solución al problema de los bucles de tierra basada en el empleo de un amplificador diferencial es muy utilizada, pero ya se ha mencionado que las tensiones en modo común pueden llegar a ser elevadas e, incluso, superiores a las que puede soportar el amplificador. En estos casos, la solución pasa por "romper los bucles" aislando galvánicamente ambos circuitos de forma que se impida la circulación de las corrientes en *modo común. Existen, fundamentalmente, tres métodos de realizar el aislamiento [29.6]* (figura *29.13*):

1. Utilizando un amplificador de aislamiento: es el método más utilizado en aplicaciones industriales cuando la señal que se envía es de tipo analógico.
2. Utilizando un optoacoplador: es la solución más económica cuando la señal que se envía es de tipo digital.
3. Utilizando un transformador: es una solución únicamente válida cuando la señal es alterna. El transformador deberá presentar una elevada tensión de aislamiento entre primario y secundario.

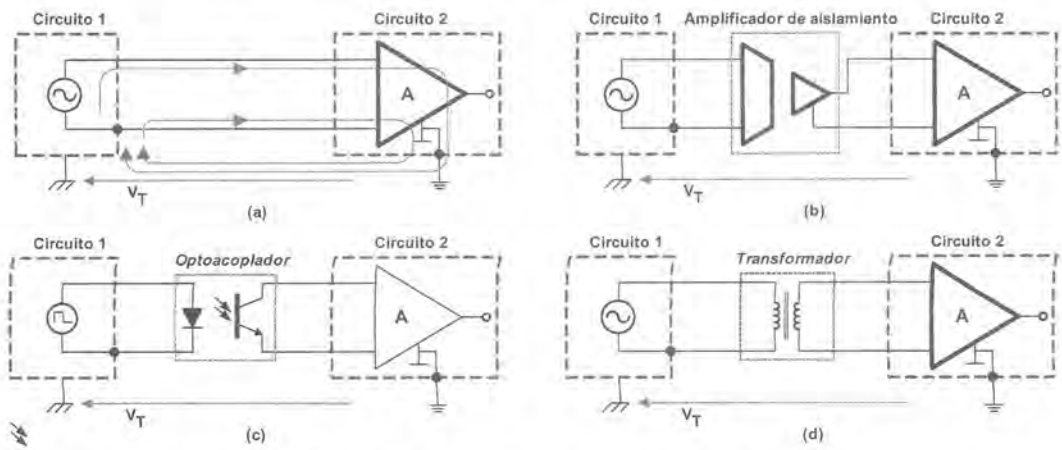


Figura 29.13. (a) Presencia de bucles debido a que los dos circuitos están puestos a tierra; (b) ruptura del bucle empleando un amplificador de aislamiento; (c) mediante un optoacoplador; (d) mediante un transformador de aislamiento.

29.1.2.2 Punto de conexión a tierra

La consideración de una señal como de alta o baja frecuencia depende de la longitud del cableado como ya se ha mencionado en el capítulo anterior. En los sistemas de instrumentación las señales son, generalmente,

de baja frecuencia aunque pueden existir subsistemas de alta frecuencia (por ejemplo, microprocesadores). Para señales de baja frecuencia, la conexión a tierra debe realizarse en un solo punto que debe ser aquel donde se referencian las masas de todos los circuitos (figura 29.14).

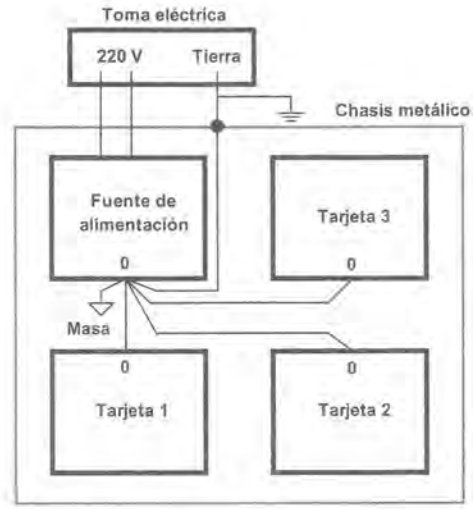


Figura 29.14. Ejemplo de conexión a tierra en un punto.

La conexión multipunto resulta necesaria cuando se trabaja con señales a alta frecuencia. En estos casos, también puede ser recomendable puestas a tierra híbridas, que se comportan como si fueran a un punto para bajas frecuencias y como multipunto para frecuencias grandes (figura 29.15).

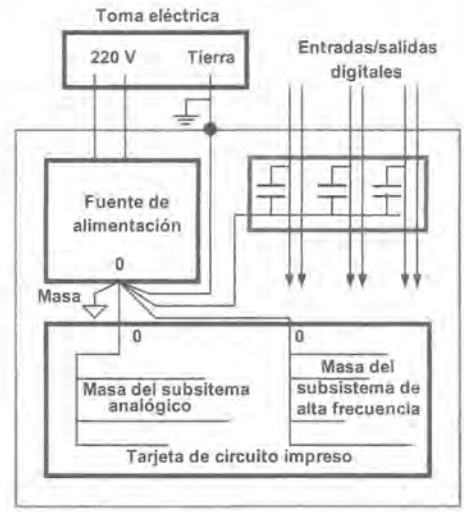


Figura 29.15. Ejemplo de conexión a tierra híbrida. Las masas de cada una de las entradas/salidas digitales de alta frecuencia se conectan a tierra a través de un condensador cuya impedancia a bajas frecuencias es muy alta por lo que la conexión se comporta como monopunto, mientras que, a altas frecuencias, la impedancia es muy baja dando lugar a una conexión efectiva con tierra (adaptado de [29.3]).

29.2 Minimización de interferencias debidas a acoplamientos inductivos

Supóngase que se conecta un sensor o un circuito se conecta a un instrumento cualquiera sin tomar precaución alguna, tal y como se muestra en la figura 29.16 (el sensor se ha representado por su equivalente de Thévenin y el instrumento, por su resistencia de entrada). Si la resistencia de salida del sensor, R_s y la del cableado son pequeñas en relación con la de entrada del instrumento, la señal que llega al instrumento es la misma que la proporcionada por el sensor. Sin embargo, en presencia de un campo magnético se generará una interferencia notable. Aparecerá una tensión de interferencia V_{int} (como se ha visto en el capítulo anterior, las interferencias magnéticas se modelan como una fuente de tensión) cuyo valor dependerá tanto del propio campo como del área encerrada por el cableado y esta puede ser notablemente elevada dependiendo de la distancia a la que se encuentren ambos circuitos.

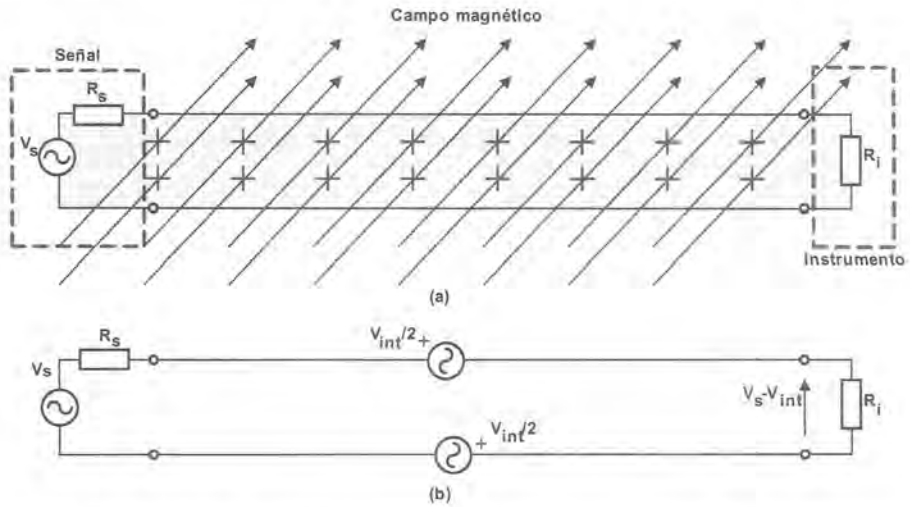


Figura 29.16. (a) Conductor paralelo sometido a un campo magnético; (b) circuito equivalente.

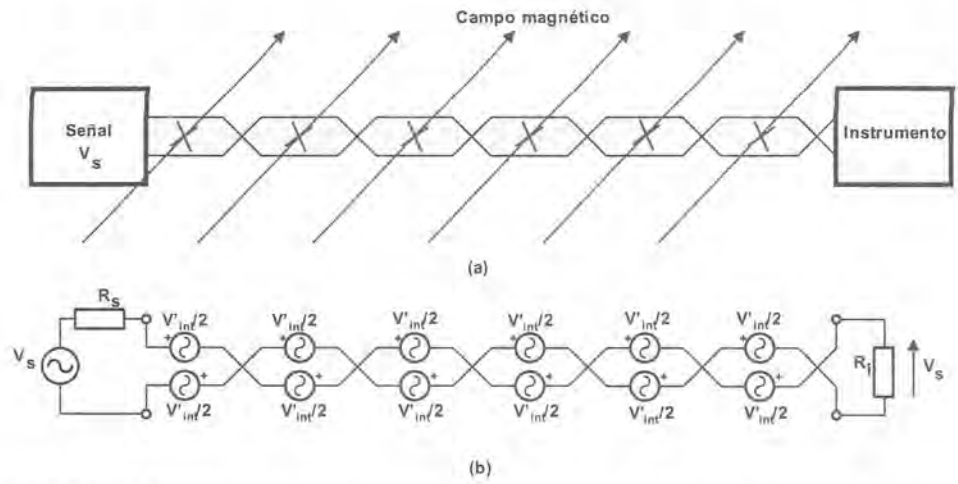


Figura 29.17. (a) Conductor trenzado sometido a un campo magnético; (b) circuito equivalente.

La interferencia disminuiría si se juntasen los conductores puesto que se reduciría la superficie encerrada, pero existe aún una solución mucho mejor: si se trenzan los conductores no sólo se reduce la superficie, sino que se generan interferencias a lo largo de los conductores cuya suma desde los extremos es nula (figura 29.17), es decir, se cancelan las interferencias.

Aunque en la práctica no se produce la anulación completa de las interferencias magnéticas debido a que los bucles no son exactamente iguales, no están sometidos a campos magnéticos idénticos, etc., su disminución es muy notable. Por este motivo, la mayoría de los sistemas de instrumentación se interconectan mediante pares trenzados cuyo coste es, además, muy reducido.

29.3 Minimización de interferencias debidas a acoplamientos capacitivos

Ya se ha mencionado en el capítulo anterior que el acoplamiento capacitivo puede reducirse empleando una pantalla que esté referenciada a la masa del circuito víctima y que rodee completamente al circuito víctima (deberá prestarse especial atención a las conexiones, puesto que en estos puntos puede resultar difícil encerrar perfectamente a los conductores).

EJERCICIO 29.1

Se desea conectar un termopar de unión aislada a un circuito de acondicionamiento. ¿Qué tipo de cable resulta adecuado?

SOLUCIÓN:

El termopar proporciona una tensión muy reducida, normalmente algunos milivoltios y su sensibilidad es pequeña (alrededor de unos 40 μV/°C, aunque depende del tipo). El acoplamiento inductivo da lugar a la aparición de tensiones de interferencia que, dados los pequeños valores de las tensiones a medir pueden ser, relativamente, muy grandes. No cabe duda alguna de que sería obligado utilizar un par trenzado para realizar la conexión.

Sin embargo, el acoplamiento capacitivo depende de la impedancia del circuito víctima, como se ha demostrado en el capítulo anterior. El termopar presenta una resistencia de salida relativamente pequeña por lo que las interferencias capacitivas serán pequeñas. Por lo tanto, no sería tan necesario que el par trenzado estuviera apantallado. De hecho, los termopares comerciales no suelen incluir pantalla.

El cable trenzado apantallado se utiliza para la transmisión de señales de baja frecuencia. Para señales de alta frecuencia son muy importantes las impedancias existentes entre los conductores y la pantalla. En estos casos se emplean cables coaxiales que están constituidos por un conductor interno, un aislante y el conductor externo que actúa como retorno de la señal a la vez que de pantalla (figura 29.18). Esta construcción permite una mejor definición de la impedancia y aunque, en principio, no proporciona buena protección frente a campos magnéticos de baja frecuencia (no hay trenzado), ofrece apantallamiento electrostático y frente a interferencias radiadas (véase apartado 29.4).

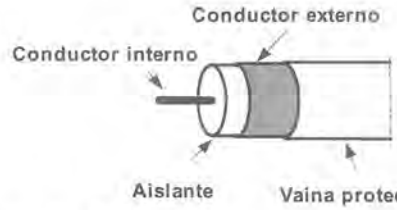


Figura 29.18. Estructura de un cable coaxial.

Conexión de las pantallas para señales de baja frecuencia

Uno de los puntos más controvertidos sobre las pantallas reside en cuál es el punto adecuado de conexión a masa. Supóngase que se pretende conectar dos sistemas, uno de los cuales no está referenciado a tierra (circuito 1) y el otro sí (circuito 2). Sabemos que las pantallas de los cables deben conectarse a la masa del circuito, pero, ¿cuál es el punto de conexión adecuado (figura 29.19a)? Básicamente, existen tres posibilidades [29.6], [29.7]:

- A la masa de ambos circuitos (figura 29.19b): en este caso, se tendrían bucles entre la pantalla y los conductores que podrían dar lugar a corrientes de interferencias por los conductores de la señal.
- A la masa del circuito no referenciado a tierra (figura 29.19c): en este caso, la diferencia de potencial existente en la tierra se acopla a través de la capacidad existente entre la pantalla y tierra dando lugar, de nuevo, a interferencias.
- A la masa del circuito referenciado a tierra (figura 29.19d): es la conexión más adecuada. Las posibles interferencias no alcanzan a los conductores de la señal.

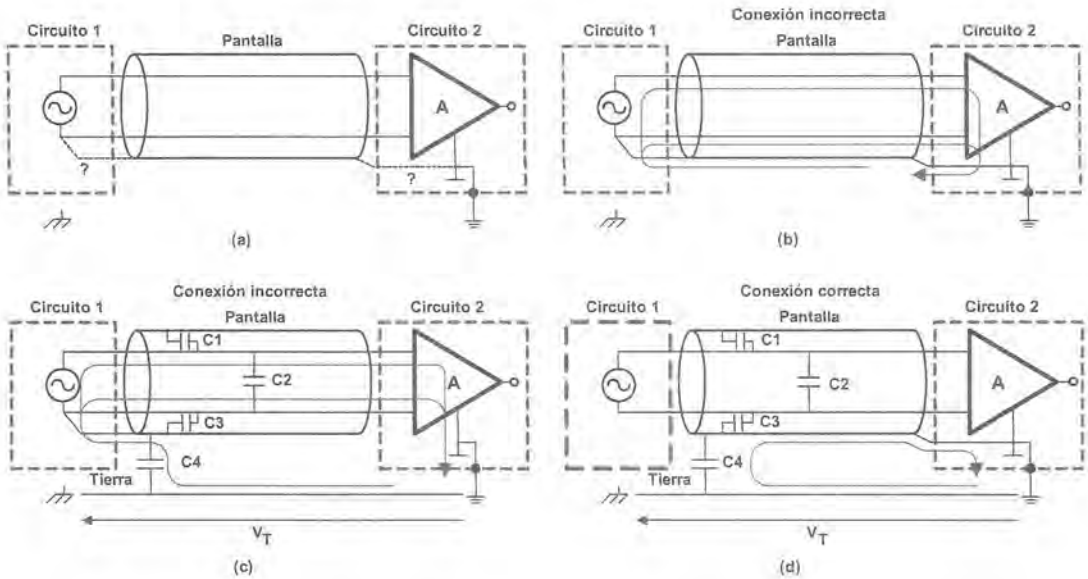


Figura 29.19. Diferentes posibilidades de conexión de la pantalla del cableado. (Para una mayor claridad no se ha dibujado el trenzado de los conductores.)

Si los dos circuitos están referenciados a tierra, aparecen bucles de tierra como se ha comentado en el apartado anterior. En este caso no está muy claro cuál resulta la mejor conexión. En general, lo más conve-

niente es conectar la pantalla a la masa de ambos circuitos puesto que parte de las interferencias originadas por los bucles derivarán por la pantalla resultando menos afectados los conductores [29.6] (figura 29.20). En algunos casos, puede resultar más conveniente efectuar conexiones híbridas (figura 29.21).

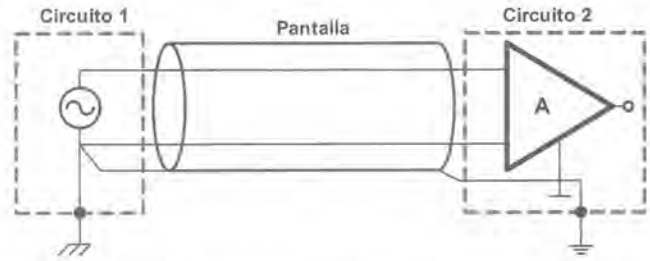


Figura 29.20. Conexión adecuada de la pantalla cuando los dos circuitos a interconectar están referenciados a tierra.

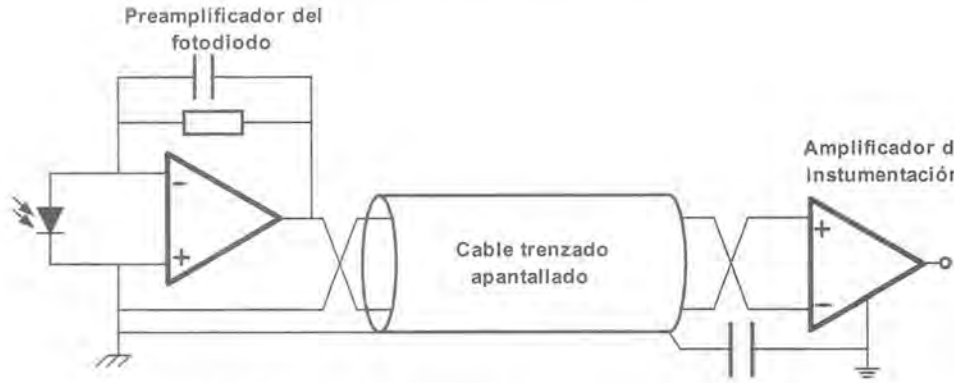


Figura 29.21. Ejemplo de conexión híbrida (tomado de [29.4]).

Merece la pena recalcar que en el estudio que se ha mostrado sobre la adecuada conexión de las pantallas se ha supuesto que las señales son de baja frecuencia. La conexión de la pantalla en un solo punto resulta muy problemática para señales de alta frecuencia puesto que se comportaría como una antena; además, las capacidades existentes entre la pantalla y tierra presentan impedancias muy pequeñas para frecuencias elevadas con lo que existe conexión aunque no sea explícita. Realmente, para altas frecuencias conviene conectar la pantalla en tantos puntos como sea posible [29.6].

No obstante, el lector puede preguntarse: ¿pero acaso no pueden estar presentes interferencias de alta frecuencia (por ejemplo, radiadas) aunque la señal a transmitir sea de baja frecuencia? La respuesta es afirmativa, pero los circuitos electrónicos deben estar diseñados para la señal a transmitir, es decir, estarán ajustados al ancho de banda necesario para la señal transmitida con lo que las interferencias de alta frecuencia son filtradas por el circuito de recepción.

29.3.1 GUARDAS ACTIVAS

Las guardas activas se utilizan cuando se desean medir señales procedentes de sensores con impedancias de salida muy elevadas (por ejemplo, sensores electroquímicos, piezoeléctricos, etc.). Supóngase, por ejemplo, que se tiene un sistema como el de la figura 29.22 donde la resistencia de salida R_s del sensor es muy eleva-

da (por ejemplo, 100 MΩ), el cableado tiene una longitud de 2 m, su capacidad es de 50 pF/m (C_c = 100 pF) y su resistencia de aislamiento es de 1000 MΩ, y el amplificador tiene una capacidad de entrada, C_a de 20 pF. Se plantearían dos problemas de importancia para realizar la medida:

1. Las capacidades y la resistencia del sensor forman un filtro paso-bajo, cuya frecuencia de corte es de 13 Hz. Es decir, el ancho de banda es muy reducido.
2. Existen fugas de corriente a través del aislante del cable[1]. Así, teniendo en cuenta el divisor resistivo que forman la resistencia de salida y la de aislamiento, la tensión que llega al amplificador se ve reducida un 10%.

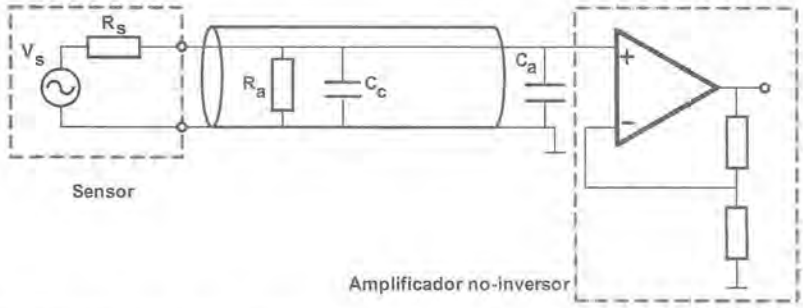


Figura 29.22. Amplificación de la señal procedente de un sensor con una elevada impedancia de salida mediante un amplificador no-inversor y conexión de los sistemas mediante cable coaxial.

Para resolver cualquiera de los problemas mencionados o ambos, se emplean las guardas activas (figura 29.23). La técnica consiste en emplear un conductor que rodee al de señal y que se mantenga exactamente al mismo potencial. De este modo no pueden existir fugas de corriente puesto que no hay diferencia de potencial ni las capacidades existentes dan lugar a filtrado alguno puesto que, dado que ni se cargan ni se descargan, no son efectivas.

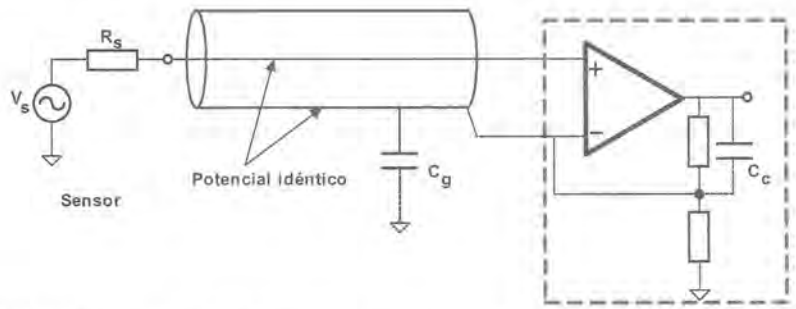


Figura 29.23. Conexión de un sensor de impedancia de salida elevada a un amplificador empleando una guarda activa.

Las guardas pueden presentar problemas de estabilidad [29.5] debidos a la existencia de una capacidad entre la misma y la masa (C_g en la figura 29.23) que puede llevar el amplificador a oscilar, sobre todo cuando la longitud del cableado es grande, por lo que suele resultar necesario emplear un condensador C_c en paralelo con la resistencia de realimentación del mismo con el fin de efectuar una compensación [29.4].

[1] Observe que en el circuito de la figura 29.22 se ha empleado cable coaxial en vez de par trenzado apantallado debido a que presenta una capacidad menor y un mayor aislamiento entre los conductores y la pantalla, con lo que las fugas son más reducidas.

Conviene no confundir una guarda con una pantalla. Pese a que el aspecto físico es similar, una guarda no da lugar, necesariamente, a un apantallamiento. Es más, en presencia de un acoplamiento capacitivo puede dar lugar a un ruido considerable. Considérese por ejemplo el caso de la figura 29.24a, donde un conductor interferente, cuyo potencial es V_r, presenta una capacidad con respecto a la guarda igual a C_r. Se puede demostrar muy fácilmente que la tensión de salida es igual a:

$$V_o = V_s \frac{R_1 + R_2}{R_2} - R_2 C_r \frac{dV_r}{dt} \tag{29.4}$$

Es decir, aparece una interferencia en forma de la derivada de la señal causante de la misma, con lo que si, además, su frecuencia es elevada, el problema puede ser muy acusado.

Para resolver este inconveniente se puede recurrir a apantallar la guarda, o bien a emplear un seguidor de tensión de baja impedancia de salida y tiempo de respuesta reducido como se muestra en la figura 29.24b.

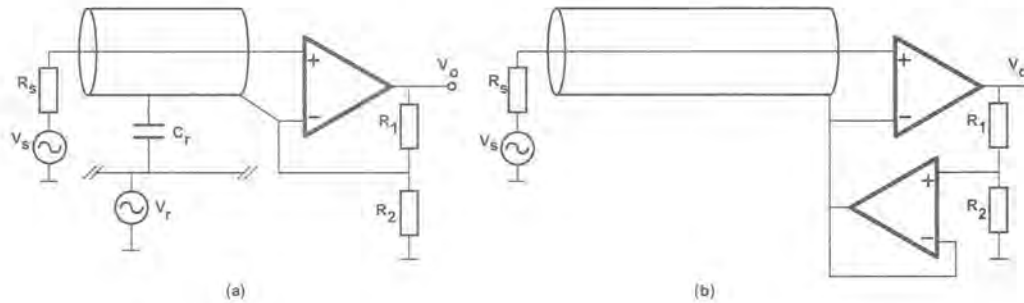


Figura 29.24. (a) Problemática de las guardas en presencia de acoplamientos capacitivos; (b) solución al problema empleando un seguidor de emisor.

En la figura 29.25 se pueden observar diferentes ejemplos de sensores y circuitos de acondicionamiento en los que puede resultar conveniente emplear guardas. Observe que en el circuito de la figura 29.25c, en la que se ha conectado una resistencia en la entrada no inversora del operacional para compensar las corrientes de polarización y la guarda está conectada a la masa aunque, según la teoría de las guardas que se ha visto, debería estar conectada a la entrada no inversora. Sin embargo, la tensión en este punto es prácticamente cero por lo que la conexión a la masa da lugar al mismo efecto y, además, se apantalla al conductor.

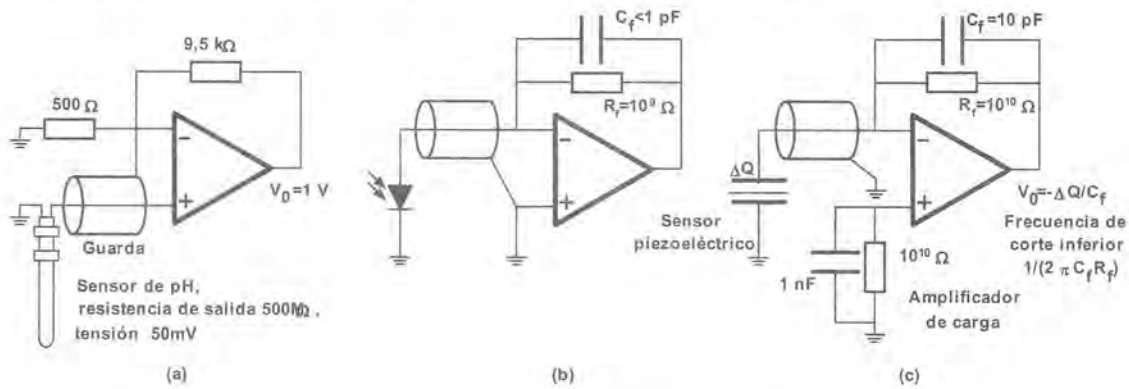


Figura 29.25. Diferentes ejemplos en los que puede resultar adecuado emplear guardas. (a) Medida de pH con un sensor electroquímico; (b) medida de niveles de luz muy bajos con un fotodiodo; (c) medida de vibraciones con un sensor piezoeléctrico (tomado de [29.2]).

29.4 Minimización de interferencias radiadas. Pantallas

Hasta el momento se han estudiado las interferencias basándose en circuitos equivalentes eléctricos, modelando el acoplamiento eléctrico por capacidades y el magnético por inductancias. Aunque es un modelo adecuado para frecuencias bajas (repetimos que la consideración de baja o alta frecuencia depende de la distancia) o mejor dicho, para campo cercano, no resulta válido para campo lejano. En este caso, el estudio hay que realizarlo a partir de la teoría general de campos electromagnéticos, lo que resulta un tanto complicado. Sin embargo, mostraremos en el presente apartado las conclusiones básicas que se obtienen, con el fin de entender los métodos de reducción de las interferencias radiadas basados en pantallas o blindajes.

Un blindaje es una superficie metálica situada entre la fuente de interferencia y la víctima. Cuando una onda electromagnética incide en el blindaje, parte de la onda se refleja en la superficie, parte se absorbe por el blindaje y parte se transmite (figura 29.26). Una parte se refleja en la parte interna del blindaje y puede seguir reflejándose sucesivamente a lo largo del mismo hasta que acaba siendo absorbida o saliendo al exterior.

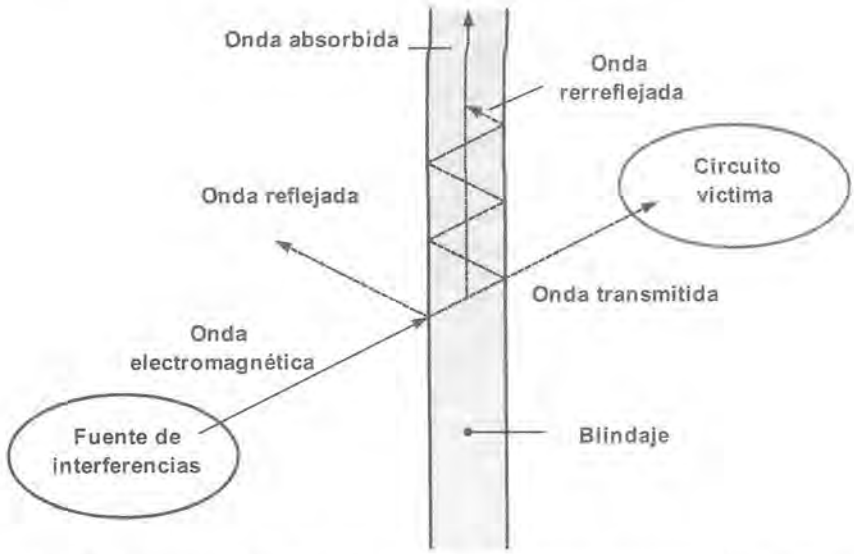


Figura 29.26. Diferentes efectos que tienen lugar cuando una onda electromagnética incide en un blindaje.

La efectividad de la pantalla, S, se expresa como la suma de las pérdidas por absorción A, reflexión R y un factor de corrección B, que tiene en cuenta las reflexiones internas [29.6]. Se expresa en decibelios:

$$S = A + R + B \ \text{(dB)} \tag{29.5}$$

Se considera que la efectividad es buena cuando está comprendida entre 60 y 90 dB y muy buena cuando supera este valor. En la tabla 29.1 se muestran los valores de cada uno de los factores en función de frecuencia de la onda, el espesor del blindaje y su conductividad y permeabilidad relativa al cobre (véase la tabla 29.2). Como se puede deducir a la vista de las fórmulas, el principal mecanismo para el apantallamiento de bajas frecuencia es la reflexión mientras que a altas frecuencias es la absorción (siempre refiriéndose al campo lejano).

	Pérdidas (dB)
Pérdidas por absorción, A	$0{,}131 \cdot t \cdot \sqrt{f \cdot \mu_r \cdot \sigma_r}$
Pérdidas por reflexión, R	$168 + 10 \cdot \log\left(\frac{1}{f} \cdot \frac{\sigma_r}{\mu_r}\right)$
Factor de corrección debido a multirreflexiones, B	Generalmente despreciable

Tabla 29.1. Pérdidas en las pantallas para campo lejano. (f: frecuencia en Hz, t: espesor del blindaje en mm, μ_r: permeabilidad relativa al cobre, σ_r: conductividad relativa al cobre) (adaptada de [29.6]).

Material	Conductividad relativa al cobre σ_r	Permeabilidad relativa al cobre* μ_r
Cobre (sólido)	1,00	1
Aluminio (blando)	0,63	1
Aluminio pulverizado con soplete oxiacetilénico	0,036	1
Zinc	0,305	1
Acero (tipo SAE 1045)	0,10	1000
Acero inoxidable (tipo 430)	0,02	500
78 Permalloy	0,108	8000

Tabla 29.2. Conductividad y permeabilidad relativa de diversos materiales (extraída de [29.5]). *La permeabilidad depende de la frecuencia y disminuye notablemente a partir de un cierto valor.

Las pantallas de los cables que, hasta el presente apartado, tan sólo se habían considerado como un mecanismo para evitar acoplamientos capacitivos, también sirven para evitar interferencias radiadas (o para evitar que se radien). Asimismo, las cajas metálicas de los equipos también forman parte esencial de la protección.

EJERCICIO 29.2

Se desea realizar una pantalla de 2 mm de espesor con alguno de los siguientes materiales: cobre, aluminio o acero. ¿Cuál es más eficaz contra las interferencias radiadas?

SOLUCIÓN:

Teniendo en cuenta las conductividades y permeabilidades relativas de los diferentes materiales (tabla 29.2) y las pérdidas por absorción y reflexión de la tabla 29.1, se pueden obtener las efectividades de los blindajes en función de la frecuencia, que aparecen representadas en la figura 29.27.

Como se puede observar, el cobre y el aluminio ofrecen un mejor apantallamiento a muy bajas frecuencias, mientras que, a frecuencias altas, es mejor el acero. De todas formas, no se ha tenido en cuenta que la permeabilidad del acero disminuye a partir de 100 kHz, de modo que a frecuencias muy altas, puede ser incluso más eficaz el cobre.

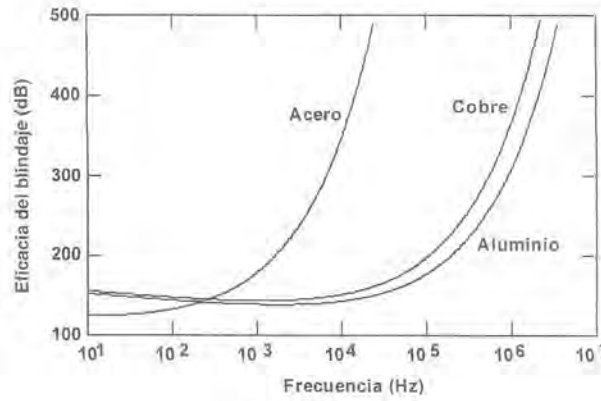


Figura 29.27

Sobre las cajas hay que tener en cuenta que los agujeros y las ranuras necesarios para la ventilación del equipo, la ubicación de botones, etc., disminuyen considerablemente la efectividad del blindaje, especialmente para frecuencias altas. Deberán tenerse en cuenta las siguientes consideraciones:

- Es mucho mejor emplear muchos agujeros pequeños que uno grande de la misma superficie.
- Deberán evitarse ranuras largas aunque sean muy estrechas.

Además, las juntas de las tapas con las cajas pueden resultar problemáticas porque, normalmente, dan lugar a discontinuidades en el blindaje ya que el contacto no suele ser muy bueno a excepción de las zonas donde se sitúan los tornillos. Para resolver el problema se pueden emplear juntas especiales conductoras ("*gaskets*") fabricadas, generalmente, con materiales elásticos, que se sitúan entre la tapa y la caja y que permiten mantener la continuidad del blindaje gracias a que se acomodan perfectamente a las irregularidades (figura 29.28).

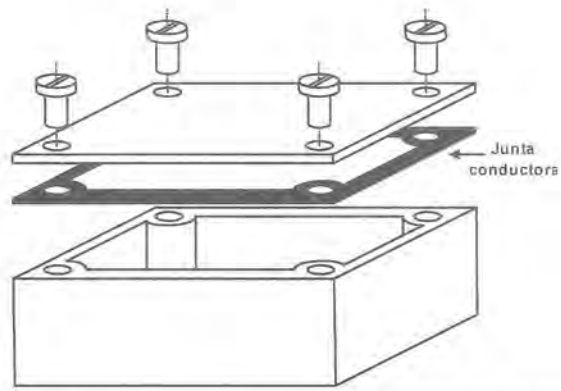


Figura 29.28. Junta conductora para mantener la continuidad del blindaje en la unión entre la tapa y la caja.

29.4.1 PANTALLAS PARA CAMPOS CERCANOS

Al analizar los métodos de reducción de interferencias para campos cercanos se ha concluido que el acoplamiento eléctrico puede reducirse rodeando al conductor víctima mediante una pantalla de alta conductividad

conectada al nivel de referencia del circuito. Si la conductividad fuera infinita, la interferencia sería teóricamente nula. El acoplamiento magnético puede reducirse rodeando a la víctima mediante un material magnético de elevada permeabilidad que conduzca el campo magnético y evite que sea alcanzada (también se puede recurrir a configuraciones que cancelen las interferencias, como el par trenzado, aunque esto no supone impedir que el campo llegue a la víctima).

Aunque el estudio es válido, hay que tener en cuenta que los materiales reales presentan conductividades y permeabilidades limitadas por lo que el modelo no explica cómo afectan estos parámetros a la eficacia de la pantalla. Para ello, resulta necesario acudir de nuevo a la teoría de campos. Los resultados se muestran en la tabla 29.3.

Como se puede observar en la tabla, la expresión para las pérdidas por absorción es idéntica a la correspondiente al campo lejano. Sin embargo, para las pérdidas por reflexión las fórmulas son diferentes y, además, hay que distinguir entre campo eléctrico y magnético (téngase en cuenta que para campo cercano la relación entre campo eléctrico y magnético no es constante).

		Pérdidas (dB)
Pérdidas por absorción, A	Campo eléctrico	$0{,}131 \cdot t \cdot \sqrt{f \cdot \mu_r \cdot \sigma_r}$
	Campo magnético	
Pérdidas por reflexión, R	Campo eléctrico	$322 + 10 \cdot \log\left(\frac{1}{f^3 d^2} \cdot \frac{\sigma_r}{\mu_r}\right)$
	Campo magnético	$14{,}6 + 10 \cdot \log\left(fd^2 \cdot \frac{\sigma_r}{\mu_r}\right)$
Factor de corrección debido a multirreflexiones, B	Campo eléctrico	$20 \cdot \log\left(1 - e^{-0{,}0303 \cdot \sqrt{f \cdot \mu_r \cdot \sigma_r}}\right)$ Generalmente, despreciable
	Campo magnético	$20 \cdot \log\left(1 - e^{-0{,}0303 \cdot \sqrt{f \cdot \mu_r \cdot \sigma_r}}\right)$ Despreciable si $A \geq 9$ dB

Tabla 29.3. Pérdidas en las pantallas para campo cercano (f: frecuencia en Hz, d: distancia en metros de la fuente de interferencia a la víctima, t: espesor del blindaje en mm, μ_r: permeabilidad relativa al cobre, σ_r: conductividad relativa al cobre) (adaptado de [29.6]).

Teniendo en cuenta las fórmulas y las permeabilidades y conductividades típicas de los materiales se puede concluir que:

- El apantallamiento frente al campo magnético mejora con la frecuencia (resulta difícil apantallar a bajas frecuencias) y el mecanismo principal es la absorción. En principio, resulta conveniente emplear un material magnético puesto que la absorción aumenta con el producto de la permeabilidad por la conductividad. Por ejemplo, este producto es 100 para el acero mientras que para el cobre es 1 (véase tabla 29.2), pero hay que tener en cuenta que la permeabilidad de los materiales magnéticos disminuye con la frecuencia por lo que, a altas frecuencias, puede resultar igual de eficaz un material no magnético.
- Para el campo eléctrico, el principal mecanismo de apantallamiento a bajas frecuencias es la reflexión para lo que interesa que la conductividad sea elevada. En altas frecuencias el apantallamiento se debe a la absorción para lo que podría resultar interesante una permeabilidad elevada salvo por el inconveniente ya mencionado de que la permeabilidad disminuye con la frecuencia.

EJERCICIO 29.3

¿Apantalla el campo magnético un material no magnético como el aluminio?

SOLUCIÓN:

Sí. Por ejemplo, supongamos que la fuente de la interferencia está a una distancia d = 1 m y que se utiliza una pantalla de aluminio de 1 mm de espesor. Teniendo en cuenta las fórmulas de la tabla 29.3 y la conductividad y permeabilidad del aluminio (tabla 29.2) se puede obtener la eficacia de la pantalla en función de la frecuencia, que se representa gráficamente en la siguiente figura (se han despreciado las multirreflexiones):

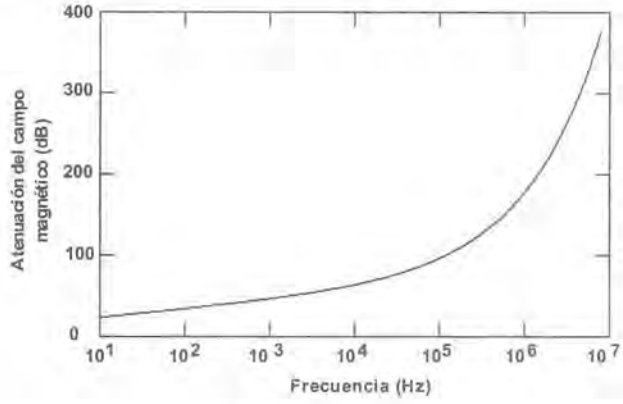


Figura 29.29

Como se puede observar, a bajas frecuencias el apantallamiento es muy deficiente (recuérdese que se consideraba que el apantallamiento era excelente a partir de 90 dB). Sin embargo, a altas frecuencias es muy eficaz.

Resumen

Deberá prestarse una especial atención al cableado de los sistemas de instrumentación para evitar interferencias.

Unas de las interferencias más comunes tienen su origen en los denominados bucles de tierra y que aparecen cuando se tienen interconectados sistemas que están conectados a tierra en puntos diferentes. Para minimizar los problemas derivados de los bucles de tierra se pueden emplear entradas diferenciales o aislar los circuitos.

Cuando se realicen conexiones a tarjetas de adquisición de datos con entradas diferenciales deberá permitirse un camino para la circulación de las corrientes de polarización del amplificador de entrada.

Para minimizar las interferencias debidas a acoplamientos inductivos deberá utilizarse par trenzado. Para minimizar las interferencias debidas a acoplamientos capacitivos deberá utilizarse cable apantallado.

Las guardas activas se utilizan cuando se desea medir señales procedentes de sensores con impedancias de salida muy elevadas. Una guarda no da lugar, necesariamente, a un apantallamiento.

Para minimizar las interferencias radiadas se recurre a la utilización de pantallas. La eficacia de una pantalla se expresa a partir de la suma de las pérdidas por absorción, reflexión y multirreflexiones internas. Las cajas metálicas de los equipos son una de las piezas clave en la protección frente a las interferencias radiadas.

Problemas propuestos

29.1 En una industria se desea medir la conductividad del líquido almacenado en un depósito metálico mediante un sensor determinado. La señal se enviará en forma de bucle de corriente a una sala de control que está situada a varias decenas de metros de distancia. ¿Qué precauciones se deberían tomar para minimizar las interferencias?

29.2 El cable para el bus IEE488 tiene 24 conductores (véase la tabla 29.4). ¿Por qué para algunas señales existe el correspondiente cable de masa mientras que para otras hay un retorno común? (Las señales DIO1...DIO8, son de datos; Las señales DAV, NFRD, NDAC son para el control de la transferencia de datos; Las señales IFC, SRQ, ATN, REN, EOI son líneas de control general del bus).

Número	Señal	Número	Señal
1	DIO1	13	DIO5
2	DIO2	14	DIO6
3	DIO3	15	DIO7
4	DIO4	16	DIO8
5	EOI	17	REN
6	DAV	18	Masa DAV
7	NFRD	19	Masa NFRD
8	NDAC	20	Masa NDAC
9	IFC	21	Masa IFC
10	SRQ	22	Masa SRQ
11	ATN	23	Masa ATN
12	Pantalla	24	Masa resto de señales

Tabla 29.4

29.3 Se desean enviar 4 señales desde una tarjeta hasta otra utilizando cable plano. ¿Cuál de las siguientes configuraciones es preferible y por qué? Si las señales 1 y 2 fueran críticas y el resto no, ¿cómo podría realizarse la interconexión?

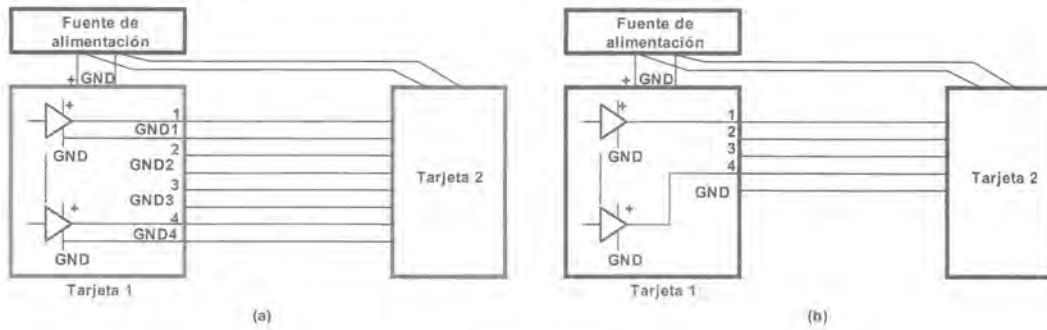


Figura 29.30

29.4 Uno de los materiales más empleados para las cajas de los equipos electrónicos es el aluminio. Discútase la eficacia de este material para apantallar el campo eléctrico, el magnético y las interferencias radiadas.

29.5 Se pretende diseñar un codificador para determinar la velocidad de giro de un gran motor asíncrono que gira a 25 Hz. Desde el punto de vista de las interferencias, ¿cuál de los siguientes materiales parece más adecuado para la envolvente del codificador: acero o aluminio?

30

Las tarjetas de circuito impreso

30.1 Aspectos generales en el diseño de circuitos electrónicos y normativa
30.2 Tarjetas de circuito impreso. Tecnologías
30.3 Consideraciones de diseño de los circuitos impresos para sistemas de instrumentación
30.4 ¿Qué debemos tener en cuenta?

30.1 Aspectos generales en el diseño de circuitos electrónicos y normativa

Cuando se termina el diseño de cualquier tipo de circuito, es decir, se ha elegido topología, se han completado todos los cálculos, simulaciones y, eventualmente, se han llevado a cabo pruebas reales, es el momento de ejecutar el montaje final. En este aspecto, la solución definitiva se llevará a cabo en la forma de una tarjeta de circuito impreso o PCB (*Printed Circuit Board*). A este proceso no son ajenos los diseños de sistemas de instrumentación electrónica y, como cualquier otro tipo de circuito, deberá implementarse sobre una PCB. En la figura 30.1 se muestra un esquema de cómo se lleva a cabo el proceso de diseño de un sistema electrónico y la documentación que debe generarse en cada paso. El proceso puede suponer regreso a etapas anteriores cuando algo falla en el proceso; debido a ello y a la participación de varias personas en el equipo, la documentación cobra una especial relevancia como vehículo de comunicación y como medio de evitar repetir una misma acción, en definitiva, como medio de optimizar el tiempo empleado en el diseño.

Cuando un diseñador poco experto completa con éxito el penúltimo paso suele pensar "¡Lo consegui![1]" y si no sale gritándolo del laboratorio es por el temor de que sus compañeros lo tomen por loco. Cree que ya ha concluido el proceso de diseño en el que, a buen seguro, ha empeñado muchas horas... Cuando adquiera un poco más de experiencia en el diseño electrónico, el diseñador sabrá que esa penúltima etapa en la que una placa de pruebas funciona correctamente es sólo el principio de un proceso en el que los problemas esperan agazapados detrás de cada componente, de cada conexión y en cada una de las pistas del circuito impreso.

[1] Algo similar dice la leyenda que gritó Arquímedes cuando cayó en la cuenta del principio de la física que lleva su nombre.

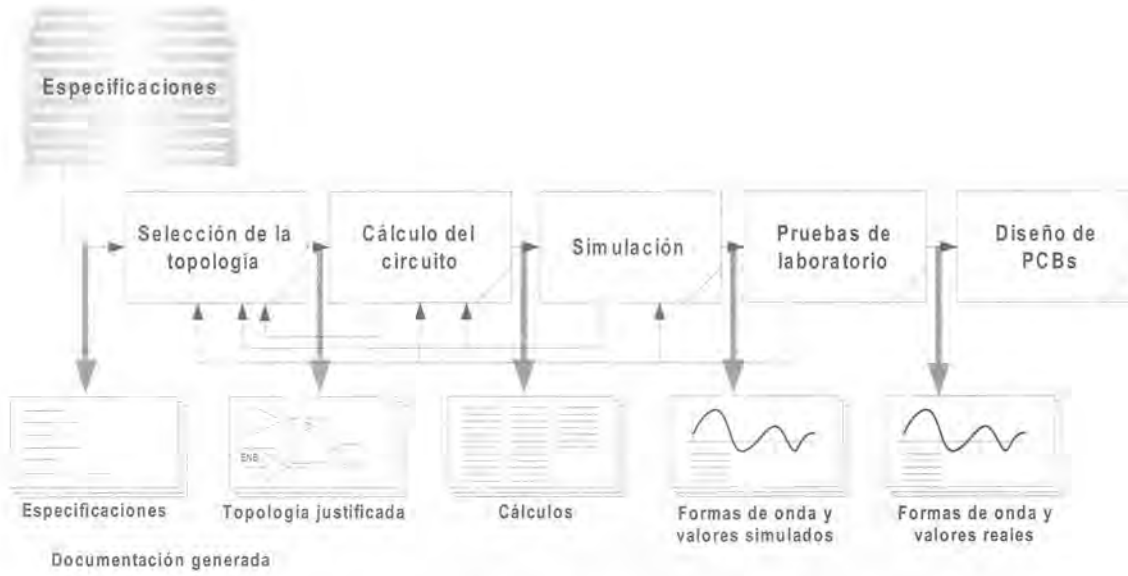


Figura 30.1. Etapas del proceso de diseño electrónico y la documentación que debe generarse en cada paso.

Aunque esto es general para todos los tipos de circuitos, cuando nos tropezamos con un circuito de instrumentación los problemas que este último paso suele presentar se incrementan fuertemente respecto a otros circuitos y la aparición de efectos negativos se multiplica en muchos casos, hasta el punto de que el correcto funcionamiento se puede ir al traste. Este aspecto cobra especial relevancia cuando se trata de un diseño que se va a producir en una serie más o menos grande en la que los aspectos estadísticos se traducirán en un número de equipos "no aptos" con el consiguiente incremento del coste de producción.

A lo largo de este capítulo proporcionaremos una pequeña guía de consideraciones sobre el diseño de tarjetas de circuito impreso que tratarán de resolver los posibles problemas derivados del último paso del diseño. No obstante, el lector debe ser consciente que aquí sólo se mencionarán los problemas y se indicarán las posibles soluciones que los eliminen o los reduzcan sin dar una demostración completa del porqué ya que esto implicaría un estudio casi siempre muy complejo y extenso que queda al margen de un libro general de instrumentación electrónica.

Además de las condiciones que impone el propio diseño habrá que tener en cuenta los aspectos de la tecnología disponible para la realización del o de los circuitos impresos y la normativa vigente:

a) La tecnología disponible no siempre es la máxima que proporcione el mercado en cada momento sino que, por lo general, quedará acotada por las limitaciones económicas que todo diseño lleva implícito. Esto podrá suponer que tecnologías que sí existan en el mercado podrían no quedar accesibles por su elevado coste para un diseño concreto. El diseñador deberá emplear la tecnología más apropiada desde los puntos de vista de prestaciones y coste e implementar sobre ella su diseño.

b) Si el diseño debe estar de acuerdo con la normativa vigente —es lo habitual y deseable—, deberíamos conocer las normas que se le aplicarían y esto puede que nos imponga alguna restricción de materiales, tamaños, forma de diseñar, etc. Más aún, las normativas, lejos de suponer un estándar universal, están limitadas geográficamente por las áreas de influencia de países y asociaciones de normalización y estandarización. Esto implica que, en algunas ocasiones, se deberán cumplir varias normas no coincidentes simultáneamente lo que acota más aún las posibilidades de diseño.

En el caso del diseño de PCBs las tecnologías disponibles y las normativas están imbricadas y, en general, las diversas posibilidades tecnológicas están recogidas y convenientemente normalizadas de tal manera que, una vez elegida una determinada solución podremos encontrar en la correspondiente norma los criterios correspondientes al diseño, a los materiales, métodos, dimensiones, ensayos, etc. Afortunadamente, el diseño de PCBs es un proceso muy establecido a nivel mundial y no hay grandes diferencias finales entre la normativa elaborada por los distintos comités de estandarización.

Una de las principales asociaciones de estandarización en el ámbito de la electricidad y la electrónica es la IEC (*International Electrotechnical Commission*)[2] a la que pertenecen la mayoría de los países desarrollados o en vías de desarrollo ya sea como miembros o como asociados [30.1]. España pertenece a través del Comité Nacional Español de la CEI de AENOR (Asociación Española de NORmalización)[3] [30.2]. Todas estas asociaciones se suelen dividir en diversos comités que se encargan de la estandarización de diversas áreas de la tecnología. En el caso de la IEC, los comités técnicos que están implicados en el desarrollo de circuitos impresos son el TC52 creado en 1956 y el TC91 creado en 1990. En la tabla 30.1 se muestra la estructura actual de estos dos comités con dos grupos de trabajo conjuntos aunque se está planteando un cambio de estructura en varios subcomités.

TC 52		TC 91	
Grupo	**Ámbito de trabajo**	**Grupo**	**Ámbito de trabajo**
WG1	Materiales laminados	WG1	Componentes de montaje superficial
WG6	Circuitos impresos	WG2	Procedimientos de ensamblado
WG10	Métodos de ensayo	WG3	Métodos de ensayo
JWG11	Transferencia de datos	WG5	Tecnología de soldadura
JWG5	Terminología		
JWG12	Diseño de PWBs (*Printed Wire Board*) y conjuntos		

Tabla 30.1. Estructura de los comités de trabajo del IEC concernientes a las PCBs.

AENOR, como miembro de IEC, ha recogido la normativa elaborada en el campo de los circuitos impresos a través del Comité 209 "Equipos electrónicos". En la tabla 30.2 se citan algunas normas importantes para el diseño de circuitos impresos.

Las anteriores normas cubren los circuitos impresos de diversas tecnologías, incluyendo los materiales empleados para su fabricación, los métodos de ensayo de circuitos y materiales e, incluso, algunas especificaciones de las estructuras más habituales que se emplean para contener muchos de los diseños de ámbito industrial como el caso de los denominados racks de 19" que se recogen en la última de las normas de la tabla 30.2.

No obstante, los diversos comités siguen trabajando y es muy previsible la revisión de las normas actuales bien sea para su mejora, bien para la adaptación a las nuevas tecnologías emergentes. Tampoco cabe desdeñar los esfuerzos de otras asociaciones ajenas a la IEC que, en algunos casos, colaboran con ella o que tratan de establecer otras normas más o menos restrictivas en ámbitos concretos desde el punto de vista geográfico o corporativo. Entre estas cabe citar:

- CECC (*CENELEC Electronics Components Committee*)
- EACEM (*European Association of Consumer Electronics Manufacturers*)
- WECC (*World Electronic Circuits Council*)

[2] También se conoce como CEI de las siglas en francés, *Commission Electrothecnique Internationale*.

[3] AENOR suministra cualquiera de las normas que edita y las de IEC. Para conseguirlas puede dirigirse a sus oficinas en la C/ Génova, 6 de Madrid o, como resulta más cómodo, a través de su página web. No se sorprenda del precio ya que este puede ser el único aspecto "normalizado" a nivel mundial por todas las asociaciones: es elevado.

- EIA (*Electronic Industries Alliance*)
- EIAJ (*Electronic Industries Association of Japan*)
- JPCA (*Japan Printed Circuit Association*)
- IPC (*Association Connecting Electronic Industries*)
- JEDEC (*Joint Electronic Devices Engineering Council*)

Norma	Título	Año	Equivalencias
UNE 20622	Código de símbolos para agujeros de circuitos impresos	1981	
UNE 20902	Técnica de los circuitos impresos. Terminología	1993	IEC 194 HD 142S3
UNE 21302-541	Vocabulario electrotécnico. Circuitos impresos	1992	IEC 50-541
EN 60097	Sistemas de reticulas para circuitos impresos	1996	IEC 97
EN 62326	Tarjetas impresas	1998	IEC 62326
UNE 20621 UNE 20621-2 UNE 20621-3 UNE 20621-4 UNE 20621-5 UNE 20621-6 UNE 20621-7 UNE 20621-7	Circuitos impresos. – Métodos de ensayo – Diseño y utilización de placas impresas – Especificación para placas impresas de simple y doble cara con agujeros no metalizados – Especificación para placas impresas de simple y doble cara con agujeros metalizados – Especificación para placas impresas multicapa – Especificación para placas impresas flexibles de simple y doble cara sin agujeros metalizados – Especificación para placas impresas flexibles de simple y doble cara con agujeros metalizados	 1980 1984 1983 1985 1985 1985 1985	 IEC 326-2 IEC 326-3 IEC 326-4 IEC 326-5 IEC 326-6 IEC 326-7 IEC 326-8
EN 123000	Especificación genérica: placas de circuitos impresos	1991	
EN 123100	Especificación intermedia: placas de circuitos impresos de simple y doble cara con agujeros no metalizados	1991	
EN 123200	Especificación intermedia: placas de circuitos impresos de simple y doble cara con agujeros para inserción de componentes	1991	
EN 123300	Especificación intermedia: placas de circuitos impresos multicapa	1991	
EN 123400	Especificación intermedia: placas de circuitos impresos flexibles sin agujero para inserción de componentes	1991	
EN 123500	Especificación intermedia: placas de circuitos impresos flexibles con taladros para inserción de componentes	1991	
EN 123600	Especificación intermedia: placas de circuitos impresos de multicapa flexorrígidas con conexiones transversales	1996	
EN 123700	Especificación intermedia: placas de circuitos impresos de doble capa flexorrígidas con conexiones transversales	1996	
EN 123800	Especificación intermedia: placas de circuitos impresos de multicapa flexibles con conexiones transversales	1996	
UNE 60249 (-1 a -20)	Materiales base para circuitos impresos. (Incluye 20 secciones para diversos materiales y métodos de ensayo y varias revisiones)	1996 a 2001	IEC 249 (-1 a -20)
EN 160100	Especificación intermedia: aprobación de la capacidad de fabricantes de conjuntos de tarjetas de circuito impreso de calidad asegurada	1998	
UNE 60297	Estructuras mecánicas para equipos electrónicos	2001	IEC 60297

Tabla 30.2. Normativa de AENOR relativa a los circuitos impresos.

En particular, resulta muy interesante la normativa de la IPC a la que hacen referencia muchas de las empresas que fabrican circuitos impresos tanto rígidos como flexibles. Estas normas son las que se muestran en la tabla 30.3, aunque la importancia para el ámbito de la instrumentación electrónica se restringe a las tres primeras fundamentalmente. La última es sólo un borrador de trabajo aunque está bastante perfilado [30.3].

Norma	Título	Fecha
IPC-2221	*Generic Standard on Printed Board Design*	1998
IPC-2222	*Rigid Printed Board Structure Design*	1998
IPC-2223	*Flexible Printed Board Structure Design*	1998
IPC-2224	*Organic, PC Card Format, Printed Board Structure Design*	1998
IPC-2225	*Organic, MCM-L, Printed Board Structure Design*	1998
IPC-2226	*High Density Interconnect (HDI) Structure Design*	2000
IPC-2227	*Organic Board Design Using Discrete Wiring*	

Tabla 30.3. Normativa de la IPC concerniente a los circuitos impresos.

Realmente, todas estas normas no son tan restrictivas como pudiéramos llegar a pensar y no constituyen un inconveniente añadido al diseño, sino que representan una práctica guía que nos ayuda a hacer que se facilite el paso de un diseño de laboratorio a un producto final por cuanto recogen las experiencias de un elevado número de profesionales del sector electrónico incluyendo a empresas, centros de diseño e investigadores. El esfuerzo de estos grupos [30.4] se traduce en unas normas que nos ayudarán a un diseño eficiente de nuestro sistema electrónico.

Pero si hablamos de un equipo de instrumentación, las cosas, una vez más, tienden a complicarse un poco más.

30.2 Tarjetas de circuito impreso. Tecnologías

La implementación de un diseño sobre circuito impreso se realiza a través del proceso de fabricación correspondiente a partir de los fotolitos generados por el software de diseño de PCBs[4] que se esté utilizando. No vamos a entrar aquí en discutir acerca este tipo de software aunque, si el lector está interesado en el tema, puede dirigirse a la información y manuales de programas como OrCAD®, Tango®, Caddy®, AutoCAD® o Protel® entre otros.

Seguramente, el lector está familiarizado ya con las tarjetas de circuito impreso y conoce, a grandes rasgos cómo son, cómo se utilizan y tiene una idea general del proceso de diseño; aquí vamos a hacer un breve recordatorio sobre los aspectos más destacados antes de entrar en las consideraciones a tener en cuenta en el caso de diseños de sistemas de instrumentación.

En general, un circuito impreso es un conjunto que está constituido por un soporte aislante que separa diversas capas de conexiones eléctricas adheridas al soporte. En la figura 30.2 se muestra un esquema de uno de estos conjuntos de circuito impreso constituido por dos capas conductoras separadas por una aislante.

[4] Existen muchos programas para la realización de tarjetas de circuito impreso que proporcionan más o menos ayuda (van desde un diseño manual hasta uno totalmente automático). Estos programas son capaces de generar la información necesaria para llevar a la práctica el diseño en sistemas de fabricación controlados por computador (CAM) o en sistemas de prototipado mucho más sencillos.

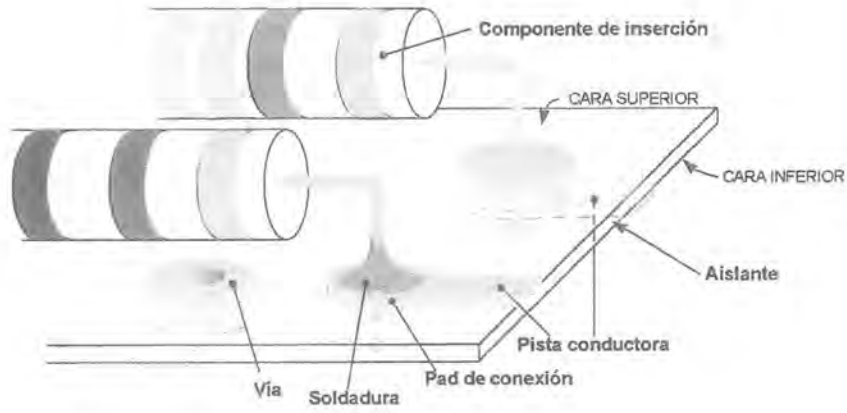


Figura 30.2. Circuito impreso con componentes de inserción.

Esta es la realización más habitual y la más tradicional en el mundo de la electrónica: la mayoría de los circuitos que forman parte de los equipos y sistemas que solemos usar están constituidos por tarjetas de circuito impreso de este tipo denominadas normalmente como circuitos impresos de doble cara y se caracterizan por un soporte rígido de algún tipo de fibra de vidrio en cuyas dos caras se disponen pistas de conexión (*tracks*), puntos de conexión de componentes (*pads*) y agujeros metalizados que conectan una cara con la otra (*vias*); por supuesto, también existen circuitos de simple cara en los que sólo hay pistas en una de las caras pero aunque su coste es algo menor que los sistemas de doble cara, la diferencia económica no suele compensar la versatilidad de trazado que proporcionan las tarjetas de doble cara.

En algunos casos especiales tales como la inclusión de sistemas electrónicos en equipos portátiles de reducidas dimensiones, se recurre a circuitos impresos flexibles capaces de doblarse con facilidad para adaptarse al poco espacio disponible; un caso muy típico de uso de circuitos impresos flexibles es el de las modernas cámaras de fotografía en las que el reducido tamaño y peso impiden el uso de circuitos impresos rígidos tradicionales. En esencia, el circuito es similar al rígido de la figura 30.2 excepto en que el soporte aislante es de un material muy flexible y de menor espesor [30.5] sobre el que se conectan dispositivos de montaje superficial (SMD o *Sourface Mounted Devices*) de muy pequeño tamaño. En la tabla 30.4 se indican algunos de los materiales más típicos empleados en la fabricación de circuitos impresos tanto rígidos como flexibles.

Tipo de PCB	Función	Materiales
Rígida	Aislante	FR-4 Fibra de vidrio
Rígida y Flexible	Conductor	Cobre (duro, recocido, depositado por *sputtering*) Aleación de cobre-berilio Aleación de cobre y níquel (70% Cu y 30% Ni) Níquel Plata
Flexible	Aislante	Kapton® Poliamida

Tabla 30.4. Materiales empleados en la construcción de tarjetas de circuito impreso.

Aunque el tipo de materiales empleados pudiera parecer secundario excepto por cuestiones mecánicas, realmente no es así ya que la presencia de conductor y de aislante en cualquier circuito impreso provoca la

aparición de parásitos de tipo resistivo e inductivo debidos a las propias pistas y de tipo capacitivo provocados por la presencia de conductores separados por un material aislante que hace el papel de dieléctrico. Por tanto, las propiedades resistivas y las características geométricas de cualquier diseño podrían tener que ser tenidas en cuenta para determinar los parásitos que forman parte del circuito.

Otros aspectos sobre los que influyen los materiales empleados son la temperatura de trabajo y la rigidez dieléctrica que, en algunos diseños pueden ser factores definitivos en la selección de los materiales de las tarjetas de circuito impreso.

La creciente complejidad de los diseños y su robustez, el trabajo a mayor frecuencia y las cada vez más restrictivas condiciones de peso y de volumen han empujado al mundo de los circuitos impresos hacia soluciones más complejas que los circuitos impresos de doble cara. Así, aparecen circuitos multicapa que están constituidos por estructuras en las que se alternan las capas destinadas a conducción y las capas aislantes según se indica en el corte vertical de una tarjeta multicapa de la figura 30.3.

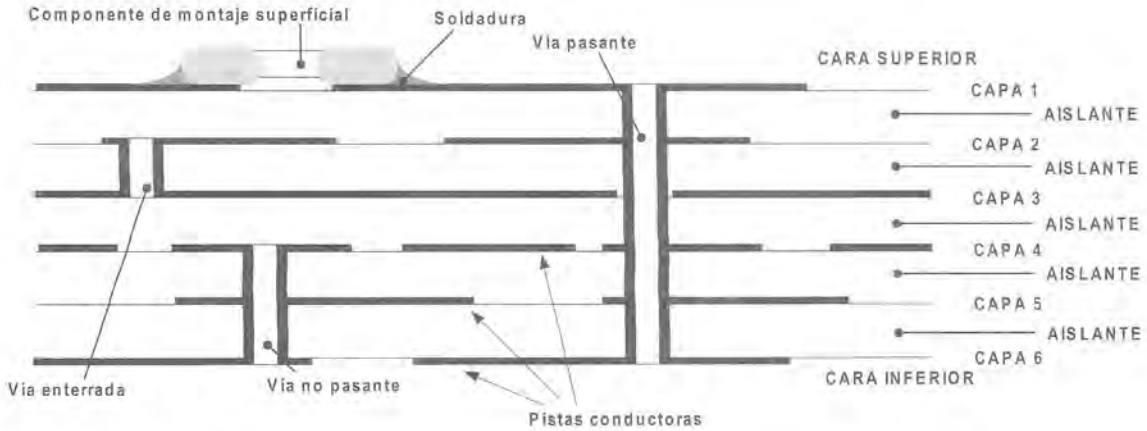


Figura 30.3. Corte vertical de una tarjeta de circuito impreso multicapa de 6 capas (la escala vertical está muy ampliada en relación con la escala horizontal).

Lógicamente estos diseños multicapa son más caros que los sencillos de doble cara pero proporcionan una gran versatilidad al diseño del circuito impreso y permiten lograr excelentes prestaciones en circuitos críticos. Por ejemplo, en circuitos digitales de alta complejidad como los computadores y los sistemas de adquisición de datos de alta velocidad, el uso de circuitos impresos multicapa es una tecnología consolidada sin la cual algunos de los equipos que nos resultan más familiares no hubieran sido posibles.

30.3 Consideraciones de diseño de los circuitos impresos para sistemas de instrumentación

Las especiales particularidades de los circuitos de instrumentación electrónica hacen que su diseño entrañe una cierta dificultad puesto que las pequeñas desviaciones de la idealidad que en otros circuitos no tienen mayor importancia aquí pueden ocasionar un grave problema o, cuando menos, contribuir a incrementar el valor de la incertidumbre. Con el diseño de los circuitos impresos para implementar estos circuitos ocurre lo mismo y el problema reaparece y no siempre a una escala menor.

Considerar el problema de los circuitos impresos para instrumentación en su conjunto supone una gran complejidad y es difícil —casi imposible— encontrar "soluciones universales" que puedan aplicarse sin dudar a cualquier diseño. Por ello, no abordaremos el problema completo, sino que nos restringiremos a proporcionar una serie de "consejos prácticos" o consideraciones que deberán ser tenidas en cuenta y que apuntarán en la dirección correcta a la hora de dar el último paso del proceso de diseño. Sin embargo, en algunos casos pueden ser irrealizables, llegar a ser contradictorias o quedar fuera de las posibilidades del diseño por cuestiones económicas. En cualquier caso, aunque no se lleguen a aplicar, es importante conocer la "sensibilidad del diseño" a diversos factores perturbadores e identificar fuentes potenciales de problemas.

Uno de los principales problemas con el que nos tropezaremos en el diseño de cualquier circuito impreso —en general en cualquier proceso de interconexión— es la presencia de bucles de cualquier tipo. Podemos preguntarnos: ¿qué es un *bucle*? La respuesta es muy sencilla: cualquier circuito que tenga al menos una malla se comporta como tal. Consideremos, por ejemplo, el circuito de la figura 30.4 en el que una señal se genera en un dispositivo A y es leída por otro dispositivo B.

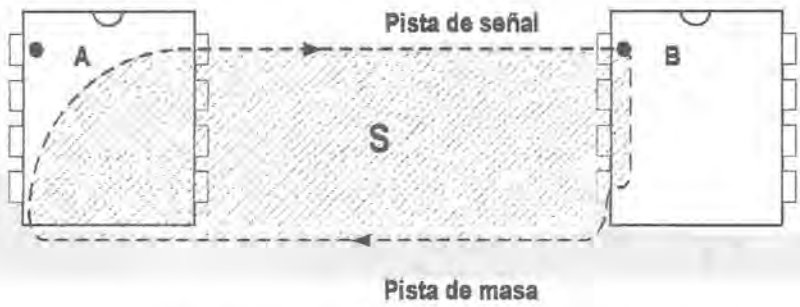


Figura 30.4. Concepto de bucle y de área de bucle, S.

El camino señalado con una línea a puntos representa la malla por la que circula la corriente entre ambos que retorna por masa; se denomina área del bucle, S a la superficie encerrada por la malla por la que circula la corriente. Pues bien, la presencia de este bucle es una fuente potencial de problemas por dos causas diferentes:

a) Cualquier campo magnético exterior podría hacer circular una corriente por ese bucle de tal manera que se añada a la de señal produciendo una pérdida de relación S/N.

b) Si se produce circulación de corriente no continua por esa malla, se generará un campo magnético que puede perturbar a otros equipos, sistemas o al propio circuito, constituyendo una potencial fuente de interferencias.

Estos dos efectos se traducen en un diseño problemático que puede no cumplir con nuestras expectativas. Por ello, sería interesante disminuir este tipo de efectos y, ya que no se puede prescindir de los bucles ya que las corrientes deben cerrarse por algún lugar configurando siempre una malla, será preciso reducir al mínimo posible el área encerrada ya que los dos efectos se reducen cuando la superficie encerrada es menor.

La mayor parte de las recomendaciones que se citarán a continuación responden, en su mayoría, a procedimientos para minimizar los efectos de los bucles a través de la reducción de las superficies. En este sentido, trataremos a continuación cómo conectar en un circuito impreso las diversas señales que tendremos: alimentaciones, masas y señales.

30.3.1 LAS ALIMENTACIONES Y LA MASA

Todos los circuitos que diseñemos incluirán una serie de componentes que precisan alimentación (*power lines*) para su funcionamiento; también hay que tener en cuenta que las señales que se generen o se reciban se referirán a un nivel de potencial determinado que suele ser el de masa (*ground*)[5], es decir, 0 V. Como quiera que los circuitos tendrán diversos componentes como éstos, se precisará distribuir a toda la tarjeta de circuito impreso las alimentaciones y la masa con el objetivo de mantener su integridad, es decir, que las alimentaciones conserven su valor de continua sin presencia de señales ajenas y que el valor de la tensión en todos los puntos de las líneas de masa sea de 0 V. Esto último es especialmente importante por cuanto todos los valores se refieren a él y, en consecuencia, cualquier cambio en el nivel de referencia afectará a todas las señales del sistema.

Para lograr la integridad de las alimentaciones y de la masa será preciso tener presentes una serie de consideraciones:

a) **Las líneas de alimentación deberán ser siempre tan gruesas como sea posible** ya que eso disminuirá los parásitos inductivos y resistivos con lo que las caídas de tensión producidas al circular corriente serán menores. Los parásitos inductivos resultan especialmente importantes cuando los consumos se producen de forma pulsada ya que puede implicar la aparición de caídas de tensión importantes que afectarían a todos los circuitos conectados al mismo punto según se indica en la figura 30.5 donde el consumo pulsante del circuito 1 (cualquier circuito digital MOS, por ejemplo) ocasiona caídas de tensión en la línea de alimentación y perturba la tensión de alimentación que recibe el circuito 2.

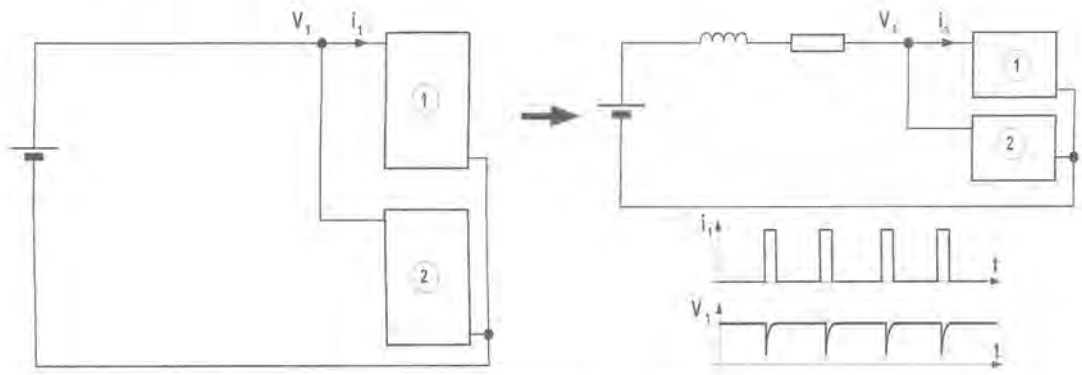


Figura 30.5. Efectos del consumo pulsante sobre las tensiones de alimentación.

El incremento de la sección del conductor (mediante incremento de la anchura de las pistas) causa una menor impedancia [30.6] y minimiza el efecto.

b) Según se indicó antes, **deberá reducirse el área encerrada por los bucles de alimentación-masa** ya que eso reducirá las corrientes de ruido inducidas por interferencias magnéticas externas. En la figura 30.6a se presenta una alimentación muy problemática por la elevada superficie presentada al *alimentar dos circuitos con un "bus de alimentación"*.

La figura 30.6b presenta una solución que mejora el comportamiento del sistema al reducir la superficie presentada por las líneas de alimentación; pero es posible una mejor solución mediante

[5] En inglés se emplea la misma palabra "ground" para tierra y masa. A veces, para tierra se emplea el término "earth-ground" pero no siempre.

una distribución en rejilla según se muestra en la figura 30.6c que reduce aún más las superficies presentadas por los bucles pero que implica el uso de dos caras de circuito impreso y ocupar un área importante sólo para las alimentaciones.

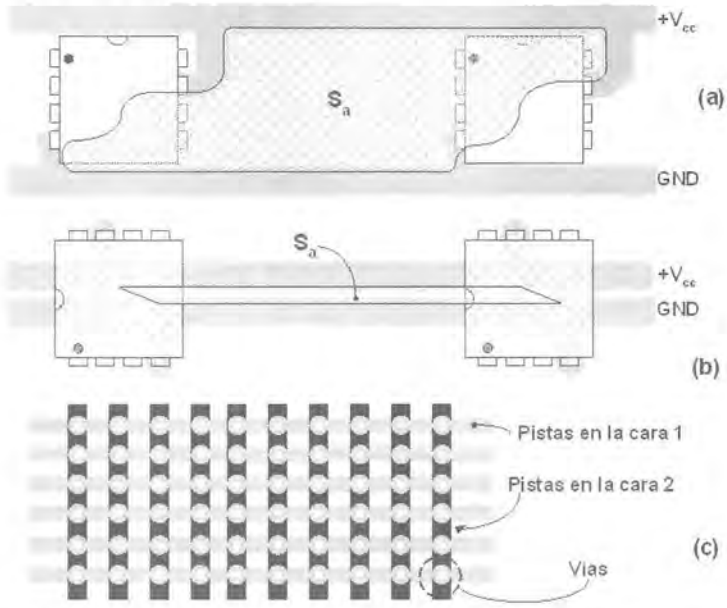


Figura 30.6. Reducción de las áreas de bucle en las pistas de alimentación: (a) elevada área de bucle debido a un mal diseño de las pistas de alimentación: (b) mejora en el trazado de las pistas de alimentación; (c) alimentación en rejilla.

c) Se deberán emplear condensadores de desacoplo de alimentación (*bypass capacitor*) situados entre alimentaciones y masa en los puntos de consumo con vistas a mantener en esos puntos a tensión constante y ayudar a proporcionar los picos de consumo de los circuitos con consumo pulsante. Sin embargo, hay que procurar que la presencia de condensadores no suponga la aparición de nuevos bucles y empeore las cosas más de lo que las mejora; en la figura 30.7a se observa una disposición incorrecta de un condensador de desacoplo que provoca la aparición de un bucle de alimentación-masa.

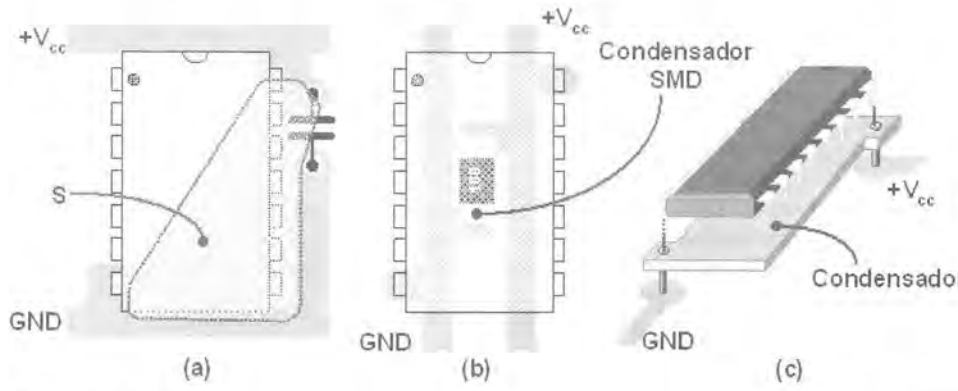


Figura 30.7. Condensadores de desacoplo de alimentación: (a) colocación incorrecta; (b) condensador SMD bajo el integrado minimizando el área de bucle, (c) condensador plano adaptado a la forma del integrado.

La figura 30.7b supone una mejora importante al emplear un condensador de montaje superficial bajo el integrado o en la otra cara y que reduce el área generada por el bucle al mínimo. Aún es posible una solución más compleja —algo más cara— como es el uso de condensadores de desacoplo de formato especial que se adaptan a los encapsulados de muchos circuitos integrados y que se colocan a modo de zócalo (o entre este y el integrado) conectando directamente la patilla (*pin*) de alimentación y el de masa (figura 30.7c).

Los condensadores que se utilicen para desacoplo deben ser capaces de responder como tales a las frecuencias de las señales que —hipotéticamente— puedan presentarse. Para ello, los valores de su resistencia serie y de su inductancia serie tendrán que ser bajos para garantizar que la caída de tensión que producen ante las señales alternas es muy pequeña y que, entonces, realmente "eliminan" el ruido de las líneas de alimentación [30.7] según se muestra en la figura 30.8.

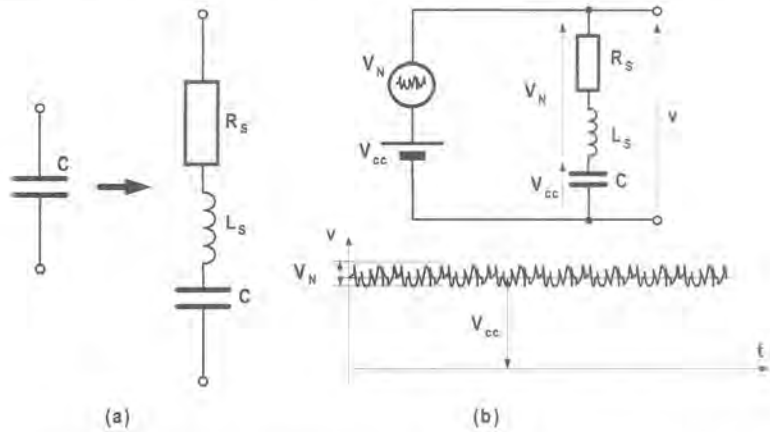


Figura 30.8. No idealidades en los condensadores: (a) modelo aproximado de un condensador real; (b) pérdida de la capacidad de filtrado de un condensador por la presencia de los parásitos resistivo e inductivo.

El valor de los condensadores de desacoplo dependerá del consumo de cada circuito pero, por lo general, casi todos los especialistas coinciden en un valor de unos 100 nF [30.8] para la mayoría de los integrados y valores mayores en las entradas de las tarjetas. En cuanto al tipo, se recomienda el metalizado con dieléctrico de teflón (MKT) por su buen compromiso entre los valores parásitos, el volumen y el precio. En este sentido, los condensadores cerámicos multicapa son una excelente solución pero su coste es alto; por otro lado, resulta totalmente contraproducente el uso de condensadores de tipo electrolítico ya que su valor de resistencia e inductancia serie es elevado.

En algunas ocasiones se pueden producir interferencias muy fuertes que pueden llevar a elevar instantáneamente la tensión en las líneas de alimentación hasta valores no sólo molestos, sino destructivos. Cuando se pretenda proteger al sistema frente a este tipo de problema se debería emplear algún tipo de dispositivo de supresión de transitorios como los TVS [30.9] que se sitúan entre alimentación y masa, justo al lado del condensador de desacoplo.

Podemos decir que, como consecuencia de lo comentado hasta aquí, el trazado de las líneas de alimentación y masa resulta complejo y que, de no hacerlo bien, las repercusiones en el funcionamiento de los circuitos pueden llegar a ser muy negativas: partes del circuito que estarán mal alimentadas, presencia de ruidos y

caídas de tensión entre diferentes puntos de masa, etc. Todos estos problemas incrementan su importancia en los casos de circuitos de alta frecuencia[6] y en circuitos que manejen señales débiles.

Existe una solución para garantizar una buena alimentación y masa en los circuitos impresos que tracemos: el uso de planos de masa y de alimentación. Un plano de masa o de alimentación es una superficie conductora que ocupa completamente una de las caras del circuito impreso tal y como se muestra en la figura 30.9.

Figura 30.9. Plano de masa ocupando una cara de una tarjeta de circuito impreso.

Como es lógico, si un sistema tiene dos alimentaciones y masa y pretendemos utilizar un plano para cada una de ellas, pensando en que precisaremos dos planos para trazar las pistas de interconexión de señales, nos conduce inmediatamente a un circuito multicapa de cinco capas (figura 30.10) lo que puede ser muy deseable pero, seguro, incrementa el coste del diseño.

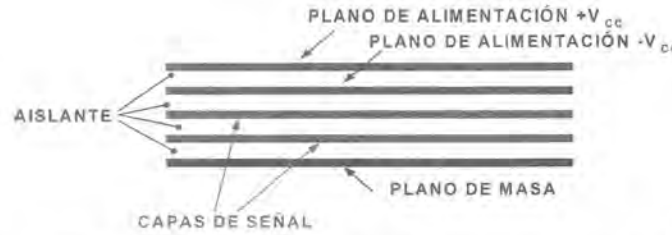


Figura 30.10. Tarjeta multicapa con dos capas de alimentación, dos de señal y un plano de masa.

Si se precisan más alimentaciones, el problema se complica más, apareciendo más capas y encareciendo considerablemente el diseño. No obstante, en diseños muy complejos y de alta frecuencia, la única opción es esta y es frecuente que los sistemas digitales de microprocesadores o DSPs incluyan tarjetas de circuito impreso de siete o más capas. La situación de las capas y el orden no debe ser cualquiera, sino que se deben seguir determinados patrones para reducir el efecto o la causa de interferencias según se indica en [30.10].

Aunque utilizar una tarjeta multicapa pueda ser prohibitivo en la mayoría de los casos, sí que puede ser interesante utilizar un plano de masa en una tarjeta de doble cara (dejando la otra cara para las pistas de señal y alimentación) o, al menos, hacerlo bajo las secciones más delicadas y sensibles del circuito según se indica en la figura 30.11.

[6] Un caso muy típico de circuito de alta frecuencia con dificultad de trazado sobre circuito impreso son las placas madre de los PCs. Muchas veces nos hemos preguntado por qué unas van mejor que otras si, aparentemente, las prestaciones de memoria, frecuencia de reloj, procesador, etc. son similares: en muchos casos, las diferencias hay que buscarlas en el trazado del circuito impreso y en la disposición de los componentes. Esto nos da una idea de la dificultad que algunos trazados entrañan.

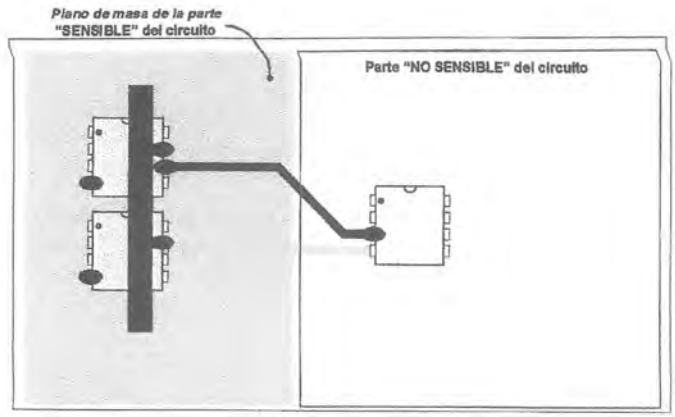


Figura 30.11. Utilización de un trozo de plano de masa bajo la parte más sensible del circuito.

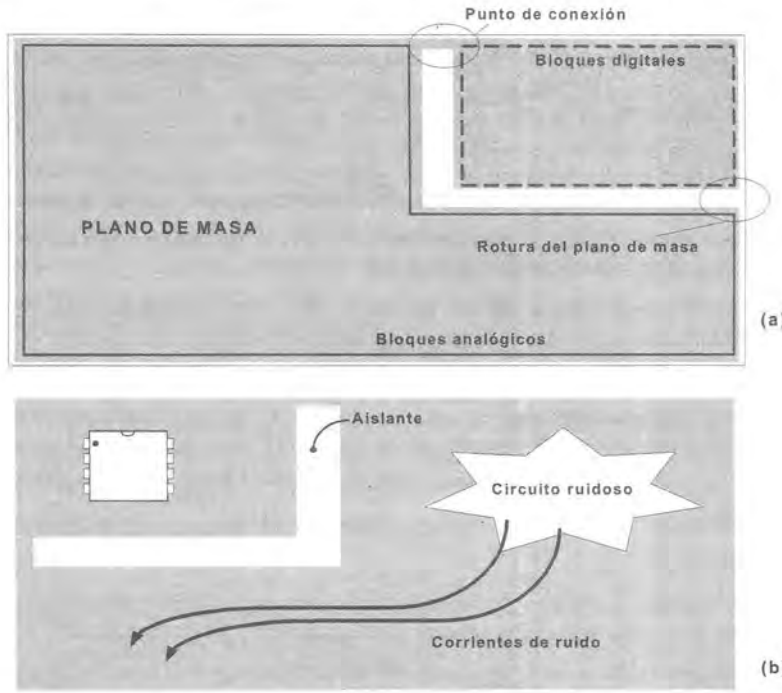


Figura 30.12. Roturas en planos de masa para mejorar el comportamiento del circuito frente a interferencias: (a) planos de masa separados para los bloques digitales y para los analógicos; (b) introducción de una zona aislante para evitar la circulación de corrientes ruidosas por zonas sensibles.

En los sistemas de instrumentación y, según hemos visto a lo largo del libro, no sólo tenemos sistemas analógicos, también aparecen bloques digitales (convertidores A/D y D/A, microprocesadores, DSPs, etc.) lo que implica que ambos sistemas aparecerán conjuntamente en muchos diseños; incluso, podría aparecer algún bloque de potencia (pequeña) en algún caso. La recomendación genérica para estos casos de diseño mixto es la **separación de los planos de masa**, dejando uno para cada función con la idea de separar los efectos que los ruidos provoquen sobre ellos. Si no hay aislamiento galvánico entre bloques —sería deseable pero también incrementa el coste—, **la conexión de todos los planos de masa debe hacerse en un único punto** [30.11] que

será la referencia global a todos los efectos. Nuevamente, esto nos conduciría a tarjetas multicapa con mayor número de capas pero, en este caso, sí que hay una solución y es dividir el plano de masa (una sola capa) en áreas sobre las que se sitúan los bloques digitales y analógicos del circuito según se indica en la figura 30.12a con lo que esta solución podría ser viable en tarjetas de doble cara. La justificación más importante es que las corrientes de ruido de cada bloque ruidoso (cada uno de ellos lo es para el otro) circulan por su masa hasta los puntos de retorno sin que afecte ni provoque caídas en el otro bloque (figura 30.12b).

En casos como este, como después justificaremos, **no se deben trazar pistas analógicas sobre la parte de plano de masa digital ni pistas digitales sobre la parte de plano de masa analógico**.

30.3.2 LAS PISTAS DE SEÑAL

El interconexionado de los distintos componentes de un diseño en circuito impreso se realiza mediante pistas de señal que llevan la información de uno a otro punto. El trazado de estas pistas resulta tan importante como el de las alimentaciones y masas y, para conseguir un buen funcionamiento, habrá que tener en cuenta una serie de criterios de diseño.

Antes de entrar en esta materia, es necesario comentar que según el circuito de que se trate, la importancia de los aspectos que mencionaremos resultará mayor o menor y que, en algunos diseños muy sencillos el trazado no resulta crítico, pudiendo funcionar "de cualquier manera". Sin embargo, como primer aspecto importante a tener en cuenta tenemos que los diseños tienen que ser más cuidadosos a medida que las señales son de mayor frecuencia y de menor potencia debido a:

a) Los conductores que conducen corrientes de alta frecuencia pueden emitir señales perturbadoras.
b) Los sistemas que trabajan a alta frecuencia pueden verse perturbados con pequeños cambios en los tiempos de propagación, instantes de cambio, etc.
c) Los sistemas que incluyen señales débiles presentan una potencia de señal reducida por lo que la relación señal/ruido puede verse muy afectada con interferencias incluso leves.

El caso de los sistemas de instrumentación era el último pero, en la actualidad, los sistemas de instrumentación pueden incluir sistemas digitales de alta frecuencia e, incluso, los bloques analógicos pueden tener que procesar señales de elevadas frecuencias por lo que, cada vez más, los sistemas de instrumentación empiezan a compartir muchos de los problemas que antes sólo eran patrimonio de los sistemas digitales.

Volviendo a las señales que podemos encontrarnos en un circuito de instrumentación, podemos agruparlas en dos grandes categorías:

1. Señales referidas a masa (*single ended*) que se caracterizan por estar representadas por la tensión de una determinada línea respecto del potencial de masa (0 V).
2. Señales diferenciales (*differential*) que quedan determinadas por la tensión que existe entre dos líneas aunque puedan tener cualquier potencial respecto de masa (tensión de modo común).

El uso mayoritario es de señales referidas a masa puesto que las de tipo diferencial exigen que el sistema que las genere tenga este tipo de salida y el que las reciba, las admita lo que es poco habitual. Por otro lado, enviar las señales como diferenciales exige el trazado de dos pistas lo que es más complejo que trazar una sola. Vamos a analizar el trazado de las pistas en ambos casos pero, teniendo en cuenta que las señales referidas a masa constituirán —sin lugar a duda— el caso más habitual que nos vamos a encontrar.

30.3.2.1 Trazado de pistas para señales referidas a masa

En la figura 30.13 aparece la interconexión de dos bloques en la que el primero envía una señal de tensión referida a masa al segundo.

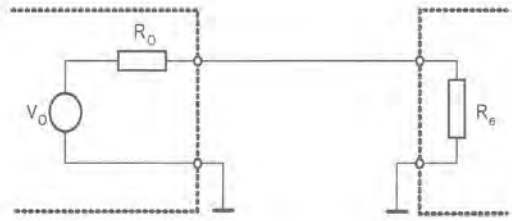


Figura 30.13. Circuito elemental para el envío de una señal entre dos bloques.

La implementación sobre circuito impreso de una conexión de este tipo no resulta, en muchos casos, tan sencilla como pudiera parecer. Veamos los problemas que se plantean y las soluciones más adecuadas (tengamos en cuenta que no siempre tendremos acceso a la solución óptima).

Bucles de señal

Como se comentó al principio del apartado, los bucles son susceptibles de recibir y producir interferencias y esto es más fácil a medida que crece el área encerrada por él. En el caso de señales referidas a masa, el retorno de la corriente se produce siempre por masa luego el área del bucle es la que determine el camino de ida y el camino de masa por el que circule la corriente. En la figura 30.14a se presenta un bucle de señal y la for-

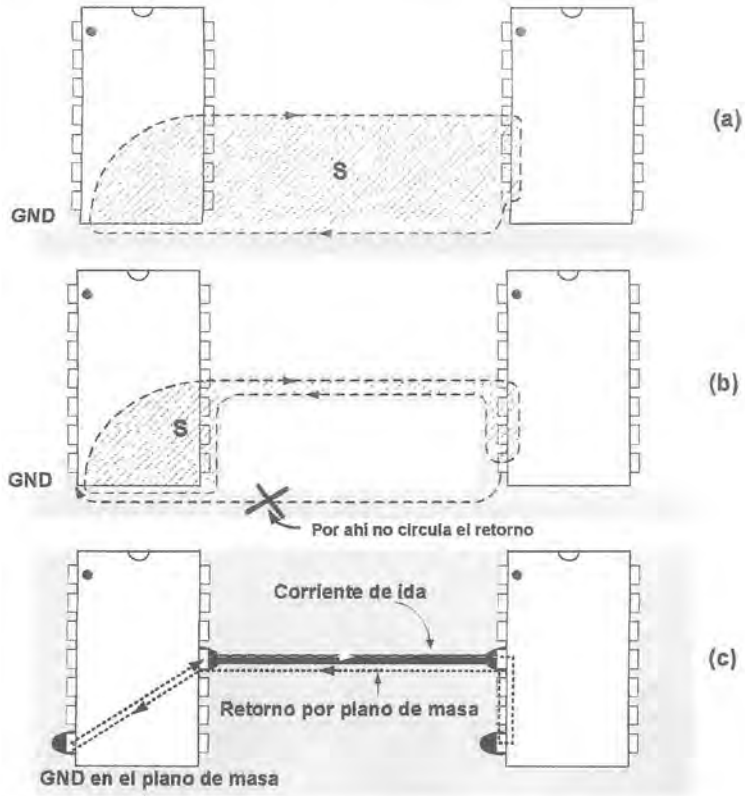


Figura 30.14. Trazado de pistas para señales referidas a masa: (a) solución inmediata que genera una gran área de bucle; (b) empleo de una pista de guarda para reducir el área de bucle; (c) el plano de masa permite el retorno más próximo reduciendo el área de bucle al mínimo.

ma de minimizar su efecto (figura 30.14b) mediante el trazado de una pista adicional de masa próxima a la pista de señal que se suele denominar *pista de guarda (guard track)*.

La corriente de retorno circula siempre por la zona más próxima al conductor de ida y no circulará por la pista general de masa sino por la pista de guarda con lo que se consigue reducir el área encerrada por el bucle de señal. El motivo para que esto sea así se debe a que el campo magnético generado por la pista de ida tiende a producir una corriente igual y de sentido contrario (igual al retorno, en el fondo) en los conductores próximos, es decir, que tiende a inducir el retorno por la guarda. Por el mismo motivo, la corriente de retorno tiende a inducir la corriente de ida en los conductores próximos (en el de ida que es el más cercano). Este acoplamiento mutuo se facilita cuanto más cerca esté la pista de ida y de la de retorno por lo que el camino más próximo será el camino de menor impedancia total.

Este fenómeno es el que justifica que, cuando se usa un plano de masa como en la figura 30.14c, se consiga el regreso por el lugar situado más cerca de la pista de ida, es decir, el camino inmediatamente debajo de ella.

El retorno de la señal por el punto más próximo hace que se tenga que ser especialmente cuidadoso con el plano de masa para evitar "rupturas" que incrementen las áreas de bucle. En la figura 30.15 se proporciona un ejemplo de diseño con la conexión de un conector de una tarjeta que es similar a lo que ocurre en un circuito integrado de inserción.

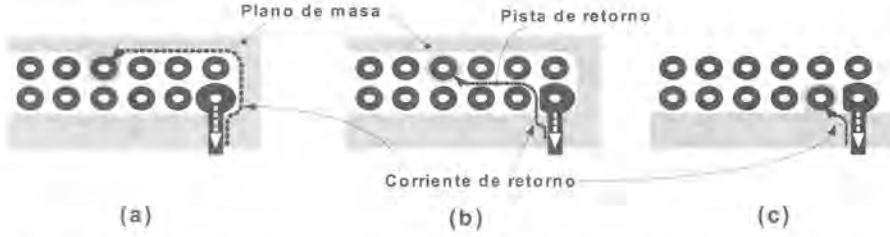


Figura 30.15. Diseño de pistas de señal en conectores compactos: (a) trazado incorrecto; (b) empleo de una pista de retorno para reducir área de bucle; (c) redefinición del conector.

En el caso (a) la señal retorna por la zona del plano de masa más próxima que —si el conector es grande— le puede obligar a dar un rodeo importante incrementando el área de bucle; en el caso (b), se ha trazado una pista adicional entre los pines del conector reduciendo considerablemente el área del bucle respecto al caso anterior lo que se traducirá en un menor riesgo de interferencias. Finalmente, si tenemos posibilidad de cambiar la definición de los pines del conector, deberíamos situar una masa al lado de cada pista de señal (o lo más próximo que sea posible) para conseguir reducir el área del bucle. Esto aumenta de importancia si se tiene un cable plano como medio de conexión en el que el área del bucle podría crecer aún más.

Exactamente lo mismo debe decirse de los circuitos integrados de inserción que "cortan" el plano de masa: resulta importante permitir el conexionado del plano entre los *pad* de sus pines para poder hacer que los caminos de retorno no tengan que rodear todo el integrado hasta llegar a la patilla de masa (figura 30.16).

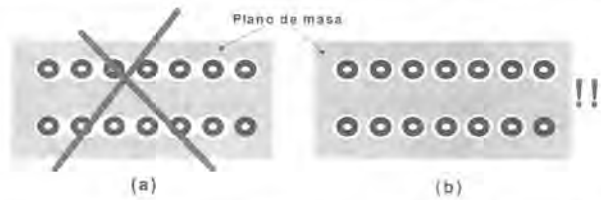


Figura 30.16. Precauciones con los circuitos integrados sobre los planos de masa: (a) mal diseño ya que el plano de masa es cortado por los *pad* de los pines del integrado; (b) diseño correcto evitando la ruptura.

La presencia de cortes en los planos de masa puede tener el mismo efecto de incrementar las áreas de los bucles de señal cuando una pista de señal pasa por encima de uno de ellos según se indica en la figura 30.17.

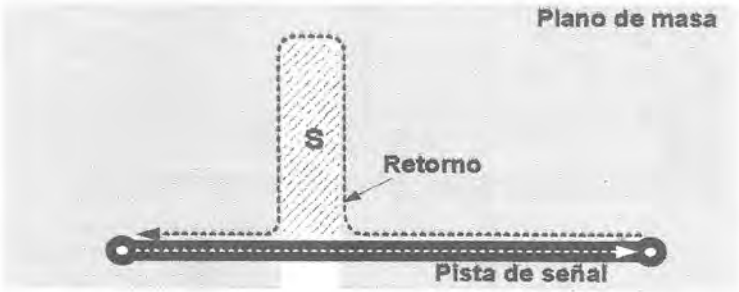


Figura 30.17. Incremento del área de bucle producido por una ruptura en un plano de masa.

El mismo efecto se tiene cuando una pista de señal de tipo analógico pasa por encima del plano de masa correspondiente a la zona digital según se indica en la figura 30.18a. En este caso, pueden ocurrir dos problemas diferentes según cuál resulte el camino de menor impedancia:

a) Que el retorno se produzca bordeando el plano de masa analógica lo que, como venimos diciendo, incrementa el área de bucle.

b) Que el retorno se produzca por el propio plano de masa digital lo que no tiene problemas de área de bucle grande pero supone un peligro de "contaminación" con ruido generado en la zona digital con lo que la separación entre plano no está cumpliendo la misión para la que fue realizada.

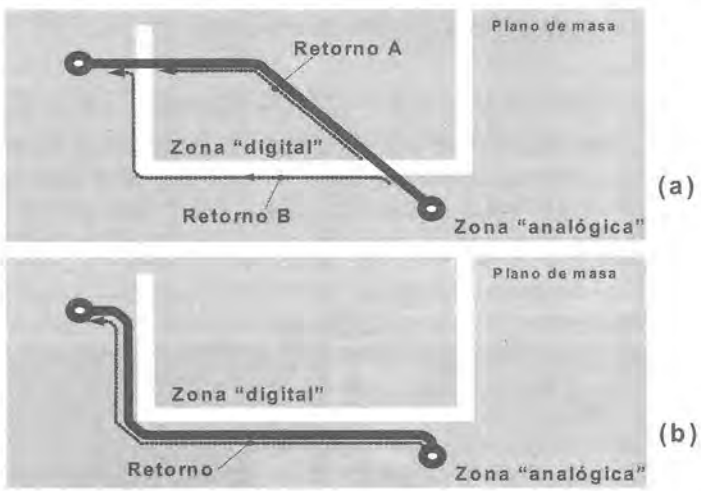


Figura 30.18. Precauciones con los cortes en los planos de masa: (a) pista de señal mal trazada; (b) pista de señal trazada para minimizar interferencias.

En la figura 30.18b se muestra el trazado correcto de la pista de señal siempre sobre el plano de masa analógico que minimiza el área de bucle y evita interferencias con la zona digital.

El problema del trazado de pistas se complica cuando existen varios planos de masa y varias capas de pistas de señal en circuitos impresos multicapa; en ese caso, el uso de vías para pasar de una a otra capa

puede complicar mucho las cosas ya que la corriente de retorno puede terminar por cerrarse por más de un plano de masa lo que podría ocasionar caminos largos y, eventualmente áreas encerradas por el bucle de señal muy grandes. En la figura 30.19 se muestra un ejemplo de lo que ocurre en un sistema multicapa en que la señal pasa de una a otra capa a través de una vía.

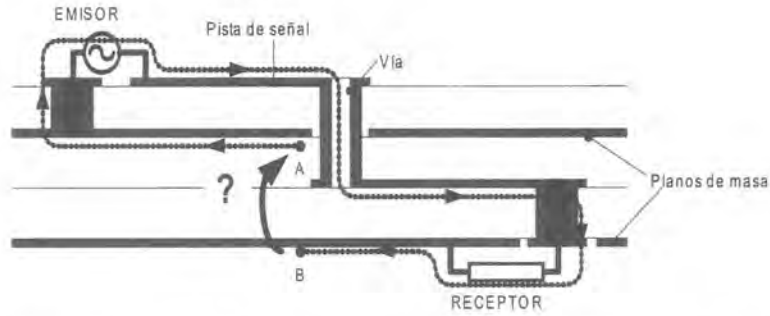


Figura 30.19. Retorno en una tarjeta multicapa. El paso desde A hasta B no está definido y puede introducir áreas de bucle muy grandes o, al menos, mayores de lo previsto.

El paso desde el punto A al punto de B de la corriente de retorno no queda muy claro por dónde ocurre y, si tiene que circular por el único punto de conexión entre planos, el área de bucle final encerrada podría ser muy grande. La mejor solución para evitar estos problemas está en no cambiar las pistas de señal de una cara a otra salvo que no quedase más remedio.

Todo lo comentado sobre el retorno de las pistas por masa puede extenderse al retorno por alimentación que también puede ocurrir; de hecho, si existen un buen número de condensadores de desacoplo entre masa y alimentación, el paso de señales de alterna de alimentación a masa y viceversa es relativamente fácil a través de cualquiera de las bajas impedancias de estos condensadores.

Conexiones en paralelo

En algunas ocasiones, una señal debe llevarse a varios puntos simultáneamente; esto puede suponer un problema serio si no se hace correctamente al aparecer bucles de señal con áreas relativamente grandes. Consideremos, por ejemplo, el caso de la figura 30.20 en la que una señal tiene que llevarse desde un punto A hasta dos puntos, B y C. Podríamos pensar que se puede trazar una pista para cada uno de los "receptores" desde el punto A (caso (a)).

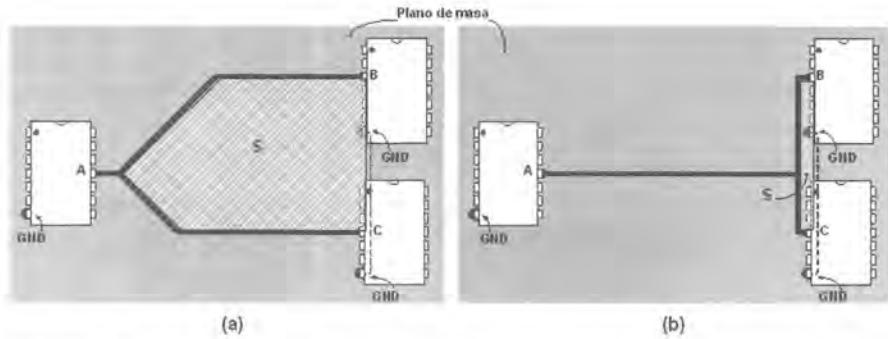


Figura 30.20. Conexión en paralelo de dos receptores: (a) trazado intuitivo en estrella que introduce un gran área de bucle; (b) trazado optimizado.

De esta forma, aunque los retornos de las respectivas señales se produzcan por el plano de masa y por el camino más próximo a las pistas de ida, la conexión entre las dos masas de los receptores, provoca la aparición de una elevada área de bucle. Por el contrario, en la figura 30.20b, se muestra un trazado más optimizado que utiliza **una única pista para llevar la señal con una bifurcación próxima a los dos receptores**, minimizando así el área de bucle.

En algunas ocasiones, no será fácil conseguir que el trazado se pueda realizar tan fácil como en este caso como, por ejemplo cuando la señal tiene que ir hacia zonas del circuito no próximas. En estos casos, puede llegar a ser interesante una redistribución de los componentes sobre el área de diseño disponible para conseguir una disposición más racional y acorde con el funcionamiento del sistema: los bloques deben disponerse en el orden determinado por la señal que se va procesando y "avanza" a lo largo de ellos, siendo poco recomendable que un bloque que vaya a tratar la señal con posterioridad se coloque antes que otro que la va a procesar antes.

El caso de señales en paralelo puede tener problemas adicionales en sistemas de alta frecuencia que precisen algún tipo de sincronización ya que habría que tener en cuenta los retardos de propagación de las señales por las pistas[7]. Para evitar problemas de sincronismo, es preciso que las señales que deban llegar "cuasi a la vez" a dos lugares, recorran distancias parecidas [30.12]. No obstante, este problema no suele ser habitual en el mundo de la instrumentación electrónica salvo en sistemas de alta frecuencia como detectores de radioactividad o medidores de luz de muy bajo nivel (contadores de fotones).

Las dimensiones de las pistas

La conexión de dos puntos entre sí rara vez se puede hacer mediante la conexión en línea recta entre ambos y la pista tendrá que sufrir algunas curvaturas. Si se realiza con una cierta precaución, esto no suele representar un problema pero tengamos en cuenta que la presencia de codos abruptos produce algunos problemas en las pistas [30.13]:

a) La aparición de codos —algunos diseñadores poco cuidadosos introducen pistas que cambian de trayectoria a 90º— resulta contraproducente por la concentración de campos en las zonas abruptas lo que contribuye a incrementar la radiación de la pista. Esto resulta particularmente importante en las líneas que transportan señales de alta frecuencia que podrían generar interferencias sobre otras partes del circuito o sobre otros equipos.

b) El cambio en el ancho de la pista en la zona de los codos provoca una modificación puntual de su impedancia característica con lo que se producirán reflexiones al considerar la pista como una línea de transmisión [30.14].

La recomendación para evitar estos problemas consiste en la realización de pistas sin cambios de dirección abruptos y sin cambios de grosor también repentinos. En la figura 30.21 se pone un ejemplo de lo que no se debe y de lo que se debe hacer [30.13].

[7] Tenemos tendencia a pensar que las señales se propagan instantáneamente pero no es así. Aunque esta propagación es muy rápida dista mucho de ser instantánea ya que queda limitada por la velocidad de la luz que no es tan rápida como creemos: un sistema trabajando con señales de 1 GHz tiene semiperiodos de 0,5 ns y una señal circulando por una pista de 10 cm tardaría algo más de 0,3 ns lo que introduciría un desfase de 120º.

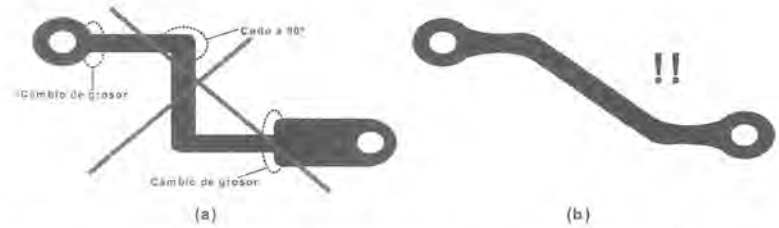


Figura 3.21. Cambios en los grosores de las pistas y ángulos a 90°: (a) mal diseño que introduce cambios en la impedancia característica y los consiguientes rebotes; (b) diseño más optimizado.

Las longitudes de las pistas también pueden llegar a ser problemáticas ya que podrían llegar a comportarse como antenas que emitan señales perturbadoras. La posibilidad de radiación sólo es real cuando las longitudes de las pistas y las longitudes de onda de las señales que circula por ellas son del mismo orden de magnitud lo que significa que, a medida que subimos la frecuencia (menor λ) las pistas deberán ser más cortas para evitar una radiación cómoda de la señal.

Diafonía

El problema de la diafonía (*crosstalk*) en los circuitos impresos tiene el mismo origen que en los cables y se produce cuando una señal que circula por una pista aparece en otra pista próxima pero no conectada con ella. Este acoplamiento se produce por las tres posibles vías: inductiva, capacitiva y resistiva según se indica en la figura 30.22.

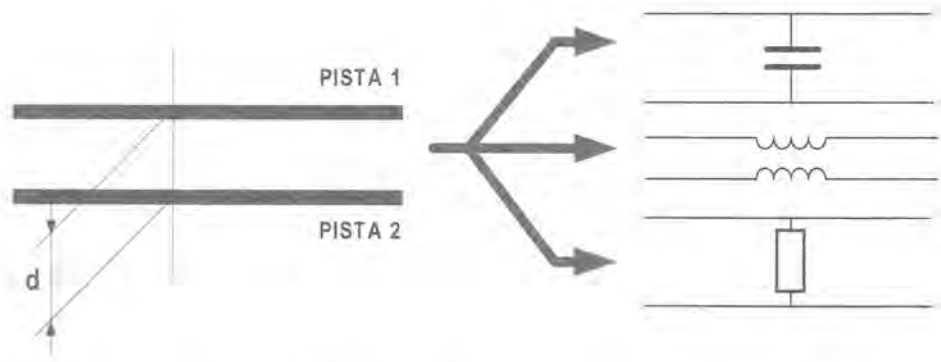


Figura 30.22. La diafonía se produce por los tres mecanismos de acoplamiento de interferencias.

La presencia del acoplamiento capacitivo es obvia puesto que las dos pistas hacen las veces de placas y el aislante que las separa, las veces de dieléctrico de un condensador plano; también lo es el inductivo ya que, en el fondo, se comportan como un par de inductores cuyos campos tendrán un cierto coeficiente de acoplamiento. Aunque suele ser menos importante, también se puede producir un acoplamiento resistivo por la no-idealidad del aislante que separa ambos conductores y, además, es más frecuente de lo que pudiera parecer: a este acoplamiento puede ayudar la frecuencia elevada que provoca pérdida de aislamiento en el dieléctrico y la presencia de partículas de polvo que podrían disminuir el valor de la resistencia efectiva entre ambos conductores.

Lo que resulta totalmente evidente es que **los problemas se agravan cuanto más cerca estén las pistas entre** sí puesto que la capacidad aumenta (la impedancia capacitiva disminuye), el coeficiente de inducción mutua se incrementa y la resistencia efectiva disminuye con lo que la "impedancia total" entre las pistas cae y es más fácil que las señales pasen de una a otra. Sin embargo, no es tan evidente lo que pasa con la longitud de las pistas que caminan juntas; como cabría suponer, el efecto de **acoplamiento crece hasta una cierta distancia en que alcanza el máximo y, a partir de ahí, se mantiene constante**. Finalmente, debido a la preponderancia de los acoplamientos inductivo y capacitivo, la frecuencia de la señal influye en la diafonía, encontrando **acoplamientos mayores a medida que sube la frecuencia**.

La forma de evitar la diafonía es, claro está, incrementar las distancias entre pistas y reducir la longitud durante la que están próximas pero eso no resulta adecuado para diseñar circuitos impresos por cuanto nos llevaría a superficies de tarjetas muy grandes; por tanto, será preciso reducir el efecto de otra manera y es mediante la colocación de planos de masa en las proximidades de las pistas. En la figura 30.23 se indica el efecto sobre la diafonía de la disposición geométrica de las pistas de señal [30.15].

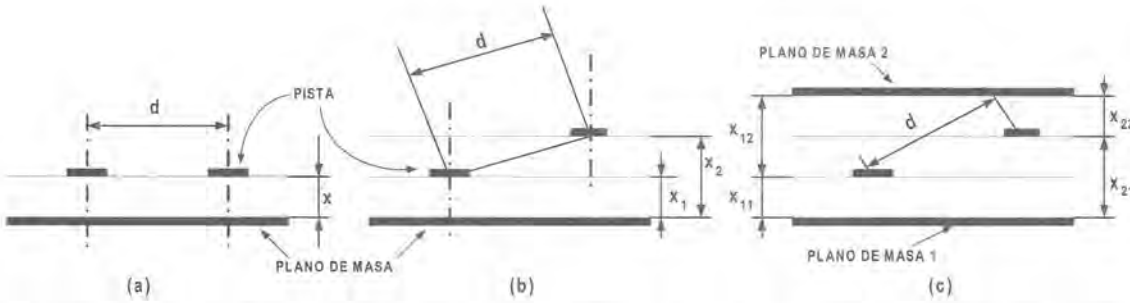


Figura 30.23. Condiciones de cálculo de los coeficientes de acoplamiento entre dos pistas de señal para una tarjeta de doble cara con un plano de masa (a), una tarjeta de tres capas con un plano de masa (b) y una tarjeta de cuatro capas con dos planos de masa (o de alimentación) (c).

En todos los casos de la figura 30.23 el coeficiente de acoplamiento es proporcional a un factor que depende sólo de la disposición geométrica de pistas y plano de masa y, en general, el acoplamiento entre las pistas crece al disminuir la distancia d que las separa y al aumentar la distancia x respecto del plano de masa. En el caso (a) que corresponde a dos pistas situadas en la misma capa (podría ser una simple tarjeta de doble cara) la expresión de ese factor geométrico resulta ser:

$$F = \frac{1}{1 + \frac{d^2}{x^2}} \tag{30.1}$$

En el caso de dos pistas situadas en distintas capas (figura 30.23b), la distancia de la expresión (30.1) debe sustituirse por la media geométrica de las distancias, es decir:

$$F = \frac{1}{1 + \frac{d^2}{x_1 x_2}} \tag{30.2}$$

Como se puede observar, con que una de las pistas esté muy cerca del plano de masa, es suficiente para disminuir el coeficiente de acoplamiento y se alcanza la peor situación cuando ambas están a la misma altura. Para minimizar más este coeficiente, se puede pensar en acercar las dos pistas al plano de masa pero, a este respecto, debemos recordar que los planos de alimentación presentan un comportamiento similar a los de masa por lo que, en tarjetas multicapa podremos tener disposiciones como la de la figura 30.23c en la que

se indican dos "planos de masa" aunque uno o ambos pueden ser planos de alimentación. Para este caso, la expresión (30.2) sigue siendo válida pero las distancias x_1 y x_2 se obtienen de las expresiones:

$$x_1 = \frac{x_{11} \cdot x_{12}}{x_{11} + x_{12}} \quad x_2 = \frac{x_{21} \cdot x_{212}}{x_{21} + x_{22}} \tag{30.3}$$

Una vez más, hay que comentar que, en el caso de no tener acceso a las tarjetas multicapa por restricciones del tipo que sea, los problemas de diafonía pueden ser paliados en tarjetas de simple o doble cara mediante la **intercalación de pistas de masa entre las pistas de señal** según se indica en la figura 30.24.

Figura 30.24. Reducción de la diafonía mediante intercalado de pistas de masa.

Como resulta evidente, el diseño de tarjetas que reduzcan el efecto de la diafonía puede resultar muy complejo y añadir restricciones graves al trazado por lo que se podría llegar a pensar en la imposibilidad de realizar un circuito impreso que no tenga este problema; sin embargo, hay que recordar que este problema no se manifiesta siempre (existe en todos los casos pero su efecto es despreciable en la mayoría) por lo que habrá que identificar las partes sensibles y evitarlo ahí nada más. Teniendo esto en cuenta, los principales problemas están producidos por pistas por las que circulen señales intensas y de alta frecuencia que pueden afectar a otras pistas que se convertirían en víctimas de las primeras. Por tanto, una buena práctica de diseño consiste en trazar las pistas de mayor nivel de señal y más frecuencia alejadas de las pistas con señales más débiles y, en el caso de que no pueda ser así, se tendrá en cuenta alguno de los métodos comentados hasta aquí para reducir efectos.

30.3.2.2 Pistas para señales diferenciales

Ya se comentó en el capítulo 29 la importancia de las señales diferenciales en la eliminación de problemas de interferencias en la transmisión de señal por la mayor facilidad para reducir el efecto del ruido en modo común. En el caso del diseño de circuitos impresos, se puede aplicar lo allí comentado aunque en una escala menor.

Por otro lado, si bien en la transmisión de señales resulta muy interesante y justificado el uso de señales diferenciales, en el caso del envío de información de uno a otro punto de un circuito impreso, generalmente carece de importancia puesto que las distancias son menores y, con ello, los perniciosos efectos se reducen. No obstante, en algunos casos, queda perfectamente justificado el uso de señales diferenciales cuando las distancias —en la escala de tamaños y frecuencias de un PCB— resulten muy grandes. En ese caso, enviaremos la información mediante dos pistas usando un emisor de salida diferencial y un receptor de entrada diferencial tal y como se muestra en la figura 30.25.

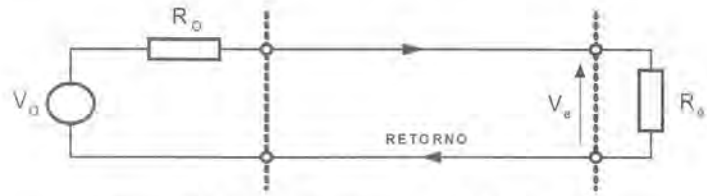


Figura 30.25. Envío de una señal diferencial.

La ventaja más importante es que podemos controlar el retorno de la señal por lo que el área de bucle queda perfectamente definida y es más difícil encontrar problemas de áreas elevadas como en el caso de las señales referidas a masa en las que la corriente retorna por masa.

Así, siempre que las dos pistas vayan muy próximas, el campo magnético de cada una de ellas tenderá a reforzar la señal en la otra con lo que:

a) Se reduce el riesgo de interferencias externas.
b) Se reduce la emisión de interferencias de las señales que circulan por ellas.

De esta manera, el uso de señales diferenciales está especialmente indicado en el caso de tener señales muy débiles que pudieran quedar expuestas a interferencias externas y en el caso de tener señales muy intensas y/o alta frecuencia que pudieran perturbar al resto del circuito.

Sin embargo, para que las señales diferenciales puedan hacer efectivas sus ventajas, resulta imprescindible hacer un buen trazado de las pistas que las llevan teniendo en cuenta una serie de aspectos que seguidamente trataremos:

Reducción del área de bucle

La reducción del área de bucle entre las pistas disminuye las posibilidades de emisión de interferencias y de aparición de señales de ruido en modo diferencial que no podrían ser eliminadas por el receptor como el ruido de modo común.

Para lograr este objetivo se trazarán las pistas muy próximas entre sí aunque, en distancias largas, el área pueda resultar inevitablemente grande. Para este caso, se recomienda realizar un proceso similar al de trenzado de los cables denominado transposición y exactamente con la misma finalidad, reducir el área final efectiva ya que se conseguiría reducir el efecto de los ruidos externos inducidos y la generación de señales de interferencia al exterior. En la figura 30.26 se indican algunos ejemplos de trazado y en el caso (c) se indica el proceso de transposición con el paso de las pistas de una a otra cara para anular los efectos de las áreas de bucle intermedias y considerar sólo las correspondientes a los extremos; este proceso complica el diseño e implica el uso de un elevado número de vías y la utilización de dos capas para una misma señal lo que puede traer más problemas por lo que su uso sólo se justifica en casos de distancias muy altas [30.16].

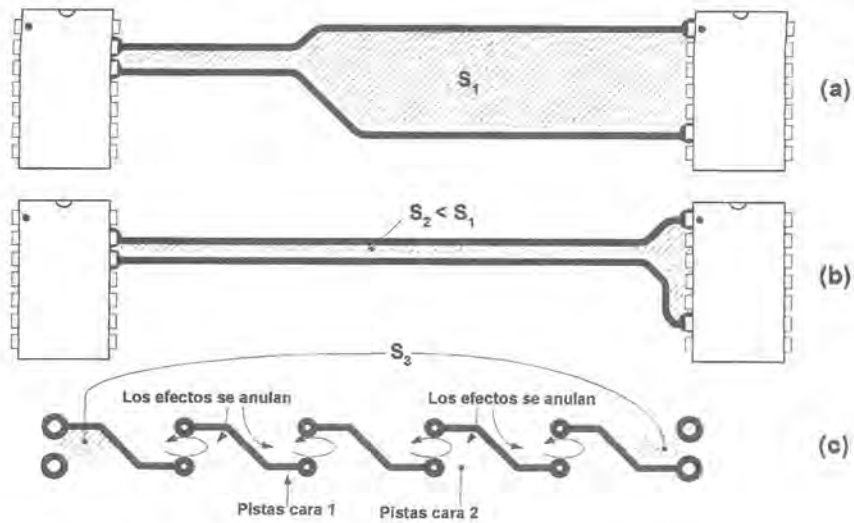


Figura 30.26. Trazado de pistas para señales diferenciales: (a) área de bucle grande; (b) el trazado próximo de las pistas permite reducir el área de bucle; (c) trenzado para mejorar la compatibilidad electromagnética.

Geometría del trazado de pistas diferenciales

La ventaja de las señales diferenciales es su autoacoplamiento que reduce los efectos de emisión exterior y que refuerza los propios, se podría perder con un mal trazado de las pistas; en este sentido, para conseguir un buen comportamiento se deberá tener en cuenta en el trazado:

a) La longitud de las pistas debe ser similar [30.17].

b) La distancia entre ellas debe mantenerse constante a lo largo de todo el camino para permitir reducir el área de bucle y para reducir el efecto de la diafonía con otras líneas del mismo circuito impreso.

En la figura 30.27a se muestra un ejemplo de mal trazado en el que las pistas se separan en algunos tramos introduciendo problemas.

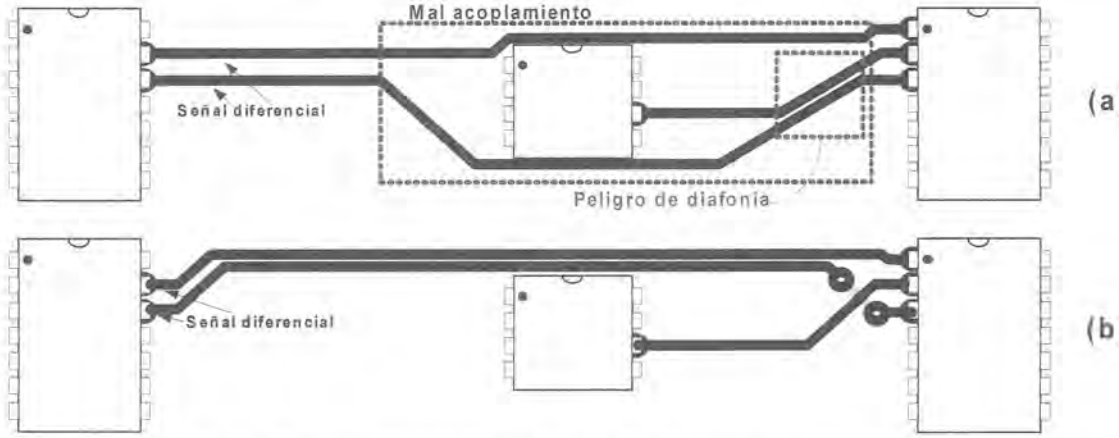


Figura 30.27. El acoplamiento y la diafonía en señales diferenciales: (a) diseño incorrecto; (b) buen acoplamiento y reducción de la diafonía.

En la figura 30.27b, sin embargo, se muestra un trazado mucho más apropiado con menos área de bucle y con menor riesgo de diafonía.

30.4 ¿Qué debemos tener en cuenta?

Si el lector ha llegado hasta aquí, puede que piense que las posibilidades de éxito sean bajas en el diseño de circuitos impresos. Cuando funciona, esperará a dar problemas a la menor oportunidad que se le presente. Realmente no es así y estas consideraciones sólo habrá que tenerlas en cuenta en algunos circuitos o en parte de ellos, no siendo de mayor trascendencia para el resto de los circuitos.

Sin embargo, la consideración de que el diseño de circuitos impresos es complejo, que se necesita tener en cuenta cómo funciona el circuito y que diseñar una tarjeta no sólo consiste en tirar los componentes de cualquier manera sobre la superficie disponible y ejecutar el auto-routing de cualquier potente programa de diseño es realmente la principal enseñanza que debemos extraer.

En estas circunstancias, cuando no funcione bien un diseño de circuito impreso correspondiente a un circuito electrónico que hayamos probado exhaustivamente cuando estaba en estado de "araña" con cables soldados entre diversos puntos y condensadores que vamos pegando a modo de parches, no debe sorpren-

dernos. Tampoco debemos buscar explicaciones extrañas como dudar de los componentes utilizados, dudar de las medidas que proporciona un equipo o, incluso replantearse la teoría. Puede que, sencillamente, nos hayamos despistado a la hora de diseñar el circuito impreso.

Las ideas que hemos presentado deben, entonces, cumplir dos misiones:

a) Constituir una guía de diseño para evitar algunos errores bastante comunes.
b) Permitir la identificación de problemas en un diseño ya realizado y, eventualmente, tomar las medidas oportunas para resolverlos.

La tarea de diseñar circuitos impresos, lejos de ser realizada por el personal de menor cualificación, suele ser llevada a cabo por expertos en el diseño de PCBs, personas con amplia experiencia y que suelen tener presentes en todo momento muchos de los factores que aquí comentamos.

Resumen

En el presente capítulo se han proporcionado una serie de ideas que deben tenerse en cuenta a la hora de afrontar la realización de una tarjeta de circuito impreso para implementar un determinado diseño electrónico. Este paso es más complejo de lo que pudiera parecer en una primera aproximación. Existe toda una normativa sobre los circuitos impresos, materiales, técnicas, pruebas, etc. y se han recogido algunas de las generadas por los más importantes organismos.

Un concepto básico es el del área de bucle que es la superficie que encierra el camino de ida y el retorno. Este valor debe ser minimizado para mejorar el auto-acoplamiento y para reducir las posibilidades de interferencias —tanto en generación como en recepción— por lo que la mayor parte de las técnicas de diseño de circuitos impresos tratan de disminuir ese valor.

Los planos de masa resultan muy importantes para conseguir que las áreas de bucle sean pequeñas pero complican las tarjetas ya que obligan a optar por placas multicapa, de mayor coste. También las alimentaciones deberían ocupar planos separados aunque eso aún incrementaría más el número de caras y, en consecuencia, el coste.

Los planos de masa pueden introducir problemas en algunos casos como cuando se tienen partes analógicas y digitales (ruidosas) en un mismo sistema. En estos casos hay que tener cuidado con los caminos de circulación de los retornos, con los cortes en los planos de masa, con los trazados de las pistas entre diversas zonas, etc. para prevenir la introducción de ruidos.

También pueden introducirse problemas desde el propio sistema como en el caso de la aparición de diafonía y en los casos en que el camino y el retorno no tienen posibilidad de buen acoplamiento. Esto se ha tratado tanto para señales referidas a masa como para señales diferenciales, indicando las diversas técnicas para conseguir minimizar los problemas potenciales: pistas de ida y retorno próximas, pistas de guarda, trenzado de pistas, separación de pistas ruidosas de las sensibles, etc.

Finalmente, no cabe despedir el capítulo sin recordar que lo que se ha contado aquí representa una tendencia, que unas veces tendrá poca importancia pero, en otras, será definitivo. Puede que nuestro diseño no funcione bien y la causa haya que buscarla en el diseño del PCB pero, habrá casos en que no podamos resolver el problema y lo expresado hasta aquí sólo nos sirva para conocer las causas.

31

Seguridad en los sistemas de instrumentación

31.1 La puesta a tierra
31.2 Seguridad en atmósferas explosivas

En el diseño de un sistema de instrumentación y de cualquier sistema, en general, resulta necesario tener en cuenta la seguridad para evitar los daños a las personas. Las normas recogen los criterios a aplicar para garantizar la seguridad con respecto a choques eléctricos, temperaturas excesivas, riesgos derivados de la propagación del fuego en un equipo, etc. Por ejemplo, las normas UNE-EN 61010 [31.5], indican los requisitos de seguridad para equipos eléctricos de medida, control y de uso en laboratorio. De todas formas, muchos equipos tienen requerimientos de seguridad particulares como los destinados a usos domésticos, los electromédicos o los destinados a atmósferas explosivas.

En este capítulo nos dedicaremos, en primer lugar, a tratar brevemente uno de los aspectos fundamentales de la seguridad desde el punto de vista eléctrico y que mayor confusión suele acarrear: la puesta a tierra. Posteriormente se tratarán los sistemas para atmósferas explosivas, centrándose en los de seguridad intrinseca, debido a su elevado interés industrial y porque son unos de los sistemas donde la normativa de seguridad es más restrictiva, afectando al diseño de la totalidad del equipo.

31.1 La puesta a tierra

La mayoría de los equipos electrónicos en general, y la práctica totalidad de los equipos de instrumentación en particular, se alimentan mediante una tensión continua de pequeño valor (entre 1,5 y 48 V) que no resulta peligrosa para las personas en ambientes secos. Sin embargo, la fuente de energía eléctrica de que se dispone es la que corresponde a la red eléctrica de 220 V eficaces (salvo en los equipos alimentados mediante baterías). Por ello se necesita un dispositivo electrónico que convierta la tensión de la red en una tensión continua y que sea capaz de proporcionar la potencia demandada por la carga. Este dispositivo es el que denomi-

namos como fuente de alimentación y lo incorporan casi todos los equipos electrónicos de instrumentación que se conectan a la red eléctrica.

La red de distribución de energía eléctrica tiene, generalmente, la disposición de la figura 31.1. Cada conjunto de usuarios (por ejemplo, un conjunto de edificios) se alimenta desde un transformador de distribución con un secundario en estrella y con tres cables de salida o fases (R, S y T). Además, existe otro cable más (neutro) conectado al punto central de la estrella.

Entre dos fases cualesquiera (R, S y T) existe una tensión 380 V eficaces, mientras que entre una de las fases y el neutro existen 220 V[1]. Pues bien, a las tomas eléctricas (por ejemplo, las de los hogares) llegan dos cables: una de las fases (cualquiera) y el neutro. Pero, además, todos conocemos que muchas tomas tienen otro terminal más: la tierra. ¿Para qué se necesita este terminal?

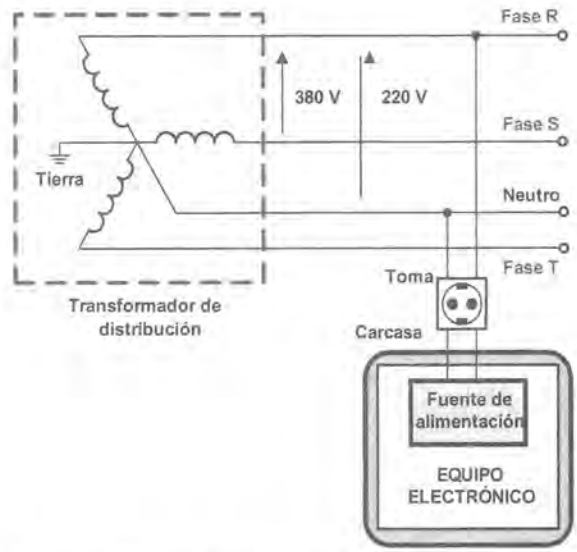


Figura 31.1. Esquema habitual de la red de distribución de energía eléctrica.

En primer lugar hay que resaltar que la tierra (o mejor dicho la Tierra, con mayúsculas) es un excelente conductor. Esta afirmación resulta muy chocante a primera vista. Si nosotros tratamos de medir con un polímetro la resistencia que existe entre dos puntos cualesquiera del suelo que pisamos, observaremos que la resistencia es, aparentemente, muy elevada. Realmente, lo que sucede es que existe una resistencia muy elevada entre las puntas del polímetro y la tierra. La resistencia que se está midiendo corresponde básicamente al contacto que se está realizando y que es muy deficiente, no a la propia tierra. Si realizáramos las medidas utilizando unas puntas muy largas, bien clavadas en el suelo, observaríamos que la resistencia es prácticamente nula. Sabemos que la resistencia de un material es proporcional a la resistividad e inversamente proporcional a la sección. La tierra puede tener una resistividad muy elevada, dependerá de las características concretas del terreno, pero la sección es enorme y de ahí que la resistencia sea muy baja.

Veamos, en primer lugar, las consecuencias que podría tener una derivación en el equipo, por ejemplo, que el cable correspondiente a la fase haga contacto con la carcasa (figura 31.2). En este caso, una persona que tocase la carcasa quedaría sometida a la tensión de 220 V; con esta tensión puede circular una corriente suficiente por el cuerpo humano como para provocar daños irreversibles.

[1] En algunos países, como Estados Unidos, los valores son diferentes.

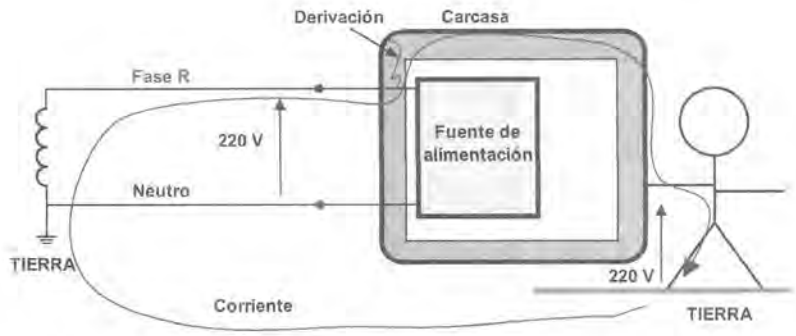


Figura 31.2. Electrocución de una persona debido a la deriva a la carcasa de un cable de alimentación de un equipo electrónico.

El problema quedaría solucionado si la carcasa estuviese conectada mediante un conductor específico a tierra. Así, la tensión entre la carcasa y tierra sería siempre nula con lo que no podría circular corriente por la persona y no cabría posibilidad de descarga alguna. Por ello, las normas obligan a conectar a tierra todas las partes metálicas de un equipo que sean accesibles por el usuario. Esta conexión se realiza mediante otro conductor (verde y amarillo) que, por ejemplo, en un edificio, va conectado a la tierra a través de unas barras especiales más o menos largas denominadas "picas de tierra" para hacer un buen contacto con el terreno.

Sin embargo, a la vista de la figura 31.2 cabe preguntarse por qué el neutro está conectado a tierra. En principio, si no se realizara esta conexión no parece que hubiera camino para la circulación de la corriente con lo que no cabría posibilidad de daño alguno. En realidad el peligro seguiría existiendo; que no haya conexión física aparente con la tierra no quiere decir que no haya conexión alguna. Entre los cables y tierra existe una resistencia (en general, muy grande) y, sobre todo, una capacidad (no se olvide que el cable es un conductor, la tierra otro, y el aire y los aislantes del cable son dieléctricos). Existe, pues, una determinada impedancia por la que puede circular la corriente. Cabría, no obstante, plantear la protección para las personas de este modo y se utiliza en algunos casos particulares [31.1]. No obstante, la mayoría de las instalaciones siguen el esquema de protección visto hasta el momento porque es el más sencillo de calcular y el más económico.

Resulta muy habitual entre los principiantes confundir los términos masa y tierra. En los circuitos electrónicos se denomina masa al conductor equipotencial de referencia respecto al que se miden el resto de los potenciales del circuito. Coincide con el cero de la alimentación y es el conductor de retorno de la mayoría de las señales activas. Para simplificar los esquemas, normalmente se dibuja con un símbolo que, según las normas [31.4], corresponde al que se muestra en las figuras 31.3b o 31.3c, reservando el símbolo de la figura 31.3d para aquellos circuitos donde existan dos masas diferentes. Sin embargo, es práctica habitual entre los diseñadores emplear para la masa el símbolo de tierra.

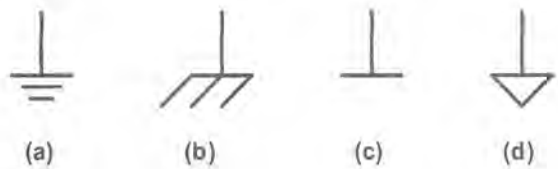


Figura 31.3. (a) Símbolo de tierra; (b) y (c) símbolos del chasis, masa; (d) símbolo de equipotencialidad.

Sin embargo, masa y tierra son conceptos completamente diferentes. Un equipo no necesita, en modo alguno, estar conectado a tierra para que funcione correctamente. Para mayor confusión, en muchas ocasiones

la masa de un circuito está conectada a tierra (véase el ejercicio 31.1) y la razón es exclusivamente la seguridad para las personas.

EJERCICIO 31.1

Se ha montado el circuito de la figura 31.4 en el que la fuente alterna es un generador de funciones y se pretende medir la tensión en el diodo mediante un osciloscopio. ¿Se puede realizar la medida?

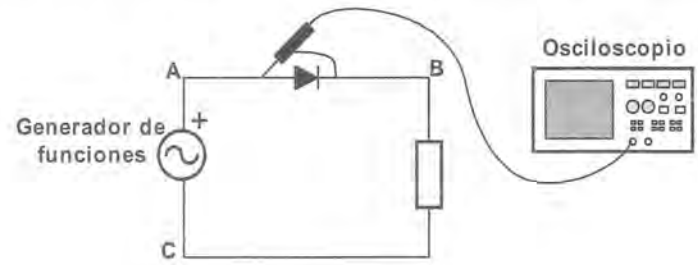


Figura 31.4

SOLUCIÓN:

No. El generador de funciones proporciona a su salida (entre el terminal positivo y masa) una tensión de bajo valor que no resulta en principio peligrosa. Sin embargo, si por un fallo en el equipo, el cable de alimentación alterna tocase al de masa, el usuario no estaría protegido. Además, el conector de salida del generador de funciones es, generalmente, de tipo BNC y, por tanto, metálico, con lo que tiene que estar puesto a tierra. Lo mismo sucede con el osciloscopio. Por lo tanto, los dos terminales negativos (las masas) de ambos equipos están unidos por tierra y si tratásemos de medir en las condiciones de la figura 31.4 estaríamos realizando un cortocircuito entre los puntos B y C.

31.2 Seguridad en atmósferas explosivas

Aquellos equipos destinados a ser utilizados en atmósferas potencialmente explosivas deben cumplir rigurosas normas para garantizar que no sean los desencadenantes de incidentes. Industrialmente es muy habitual la presencia o posible presencia de atmósferas peligrosas tanto por el propio objeto de la industria como por la necesidad de utilizar gases inflamables en muchos procesos. Básicamente, existen dos mecanismos por los que un circuito eléctrico puede producir la inflamación de un gas: por chispas o arcos eléctricos y porque la temperatura en alguna zona del circuito sea suficientemente elevada. En términos generales, los equipos que presenten algún tipo de protección están marcados con un símbolo distintivo (figura 31.5).

Figura 31.5. Símbolo de los equipos que presentan protección ante atmósferas explosivas.

Para determinar los requerimientos que deben satisfacer los equipos que se empleen en estos lugares, las normas [31.2] distinguen entre equipos de grupo I y II, y dentro de ellos se realizan diferentes clasificaciones en función, principalmente, de la propia peligrosidad del ambiente:

Equipos del grupo I: equipos destinados a minas en los que puede haber presencia de grisú[2] o polvos combustibles. Se distinguen dos categorías:

- Categoría M1: son equipos que deben de funcionar en presencia de una atmósfera explosiva. Deben tener un nivel de protección muy alto, de forma que en funcionamiento normal no pueden inflamar la atmósfera presente ni aunque se produzcan uno o incluso dos fallos en el equipo.
- Categoría M2: son equipos que deben tener un nivel de protección alto de forma que en funcionamiento normal no puedan inflamar la atmósfera o cuando se manipulan de forma descuidada. Serán retirados cuando se detecta la presencia de una atmósfera explosiva.

Para garantizar que estos equipos no son causa de la inflamación debido a temperaturas elevadas, las normas establecen que la temperatura máxima del punto más caliente no debe sobrepasar los 150°C si sobre la superficie puede depositarse una capa de polvo de carbón, o 450°C si no es probable la formación de capas de carbón.

Equipos del grupo II: destinados a atmósferas explosivas que no sean minas de carbón. Mientras que en los equipos de grupo I la peligrosidad se debe fundamentalmente a la presencia del grisú (metano), en los de grupo II los gases presentes pueden ser cualesquiera, por lo que se realiza una primera clasificación de los gases según su riesgo de inflamación en causa de chispa o arco eléctrico:

- Categoría A: comprende, entre otros, a hidrocarburos como el metano, propano, etc.
- Categoría B: uno de los gases más representativos de esta categoría es el etileno, pero incluye muchos más, como el metilacetileno, ácido cianhídrico, etc.
- Categoría C: incluye el hidrógeno, el acetileno y el bisulfuro de carbono.

Un gas de categoría C inflama más fácilmente a causa de una chispa que uno de categoría B o A, por lo que los equipos destinados a ser utilizados en presencia de estos gases deberán de cumplir mayores exigencias.

Además, atendiendo al riesgo de inflamación debido a temperaturas superficiales elevadas en los equipos, el material eléctrico de grupo II se clasifica del siguiente modo:

- T1: temperatura máxima 450°C.
- T2: temperatura máxima 300°C.
- T3: temperatura máxima 200°C.
- T4: temperatura máxima 135°C.
- T5: temperatura máxima 100°C.
- T6: temperatura máxima 85°C.

Por otra parte, no cabe pedir los mismos requerimientos a un equipo que vaya a ser utilizado donde la presencia del gas peligroso sea continua que si tan sólo es esporádica. Por ello se realiza la siguiente clasificación de las áreas:

- Zona 0: aquella en la que la atmósfera peligrosa está presente de manera permanente o durante largos periodos de tiempo. Los equipos destinados a estas zonas deben tener un nivel de protección muy alto de forma que no produzcan la inflamación ni en funcionamiento normal ni aunque se produzca uno o, incluso, dos fallos en el propio equipo.

[2] El grisú es, en realidad, metano y es producido, sobre todo, en las minas de carbón.

- Zona 1: aquella donde es probable la presencia de una atmósfera explosiva en condiciones normales. Los equipos destinados a estas zonas deben tener un nivel de protección alto, garantizando que no provoquen la inflamación ni en funcionamiento normal ni aunque se produzca un fallo.
- Zona 2: es improbable la presencia de atmósferas explosivas salvo que se produzca una avería o durante breves periodos de tiempo. Las únicas exigencias para los equipos es que no produzcan la inflamación en funcionamiento normal.

31.2.1 MÉTODOS DE PROTECCIÓN

Los mecanismos por los que se puede lograr que un equipo sea seguro son variados:

- Por inmersión en aceite (símbolo Ex o, norma EN50015): por supuesto, el aceite que debe ser menos inflamable que la atmósfera explosiva. Se utiliza, por ejemplo, para algunos transformadores.
- Por sobrepresión interna (símbolo Ex p, norma EN50016): la envolvente del equipo contiene un gas no inflamable a mayor presión que el exterior, con lo que el gas peligroso no puede alcanzar al equipo.
- Por relleno pulverulento (símbolo Ex q, norma EN50017): se rellena la envolvente con una arena que impide que un arco interno provoque la inflamación del gas exterior.
- Por envolvente antideflagrante (símbolo Ex d, norma EN50018): la envoltura externa impide que una explosión interna llegue a producir la inflamación del gas del exterior.
- Por seguridad aumentada (símbolo Ex e, norma EN50019): se utiliza con aparatos que normalmente no producen chispas ni superficies calientes en funcionamiento normal. La protección se logra, por ejemplo, mediante la utilización de aislantes mejores de lo habitual, cierres difíciles de desmontar, etcétera.
- Por seguridad intrínseca (símbolo Ex i, norma EN50020): se garantiza que la energía eléctrica en un circuito es suficientemente baja como para que no se produzca la inflamación por chispa, así como que las temperaturas en los equipos son menores que las de inflamación de los gases presentes. Se distinguen dos subcategorías: ia, ib. Los equipos de categoría ia no deben de producir la inflamación en funcionamiento normal ni aunque tengan lugar dos fallos. Las exigencias de los de categoría ib son menores, tan sólo se requiere que no produzcan la inflamación ante un fallo y, por supuesto, en funcionamiento normal.
- Por encapsulado (símbolo Ex m, norma EN50028): el aparato está encerrado en una resina que impide que las chispas o puntos calientes entren en contacto con la atmósfera explosiva.

El grado de protección que proporcionan estos métodos no es igual, por lo que para las zonas de tipo 0 sólo se permite la protección ia, que es la más segura. Para las zonas de tipo 1, es válido cualquiera de los métodos anteriores (por supuesto, también el ia).

EJERCICIO 31.2

Se dispone de un determinado equipo para la utilización en atmósferas explosivas que viene etiquetado con el siguiente código: EEx ia IIA T4.

¿Cuál es el significado de estos códigos?

SOLUCIÓN:

EEx: indica que es un equipo destinado a ser usado en atmósferas explosivas.

ia: indica que el método de protección es de seguridad intrínseca ia. Por lo tanto, incluso aunque se produzcan dos fallos en el equipo, seguirá siendo seguro. Sería adecuado para ser usado en una zona de tipo 0 (por supuesto, también para una zona de tipo 1 ó 2).

IIA: se trata de un equipo de grupo II para gases que entran dentro de la categoría A (por ejemplo, el etileno). No puede utilizarse con gases del grupo B o C.

T4: la temperatura superficial máxima es 135ºC. Los gases presentes deberán de tener una temperatura de inflamación mayor.

31.2.2 SEGURIDAD INTRÍNSECA

El método de protección de seguridad intrínseca (ia o ib) es el que tiene una relación más directa con el equipo eléctrico o electrónico. Además, de todos los posibles métodos para garantizar la seguridad de un equipo destinado a atmósferas peligrosas, el más exigente es el que corresponde a la seguridad intrínseca de tipo ia, que es el único válido para todas las posibles zonas.

La filosofía de la seguridad intrínseca consiste en que la energía que maneja el circuito sea inferior a la que se necesitaría para producir la inflamación del gas debido a chispas y en que las temperaturas de los componentes son suficientemente reducidas. Esto supone que resulta necesario calcular las resistencias, los condensadores, los semiconductores y, en definitiva, todos los elementos del circuito para que ninguno de ellos sea causa de un accidente. Las normas aplicables son, fundamentalmente, las UNE-EN50014, 50020 y 50039.

31.2.2.1 Riesgo de inflamación debido a chispas

Para el cálculo de un circuito de modo que no exista riesgo de chispas que sean capaces de inflamar el gas hay que tener en cuenta, en primer lugar, que las atmósferas peligrosas están clasificadas en grupo I, IIA, IIB y IIC, como ya se ha mencionado, y que las exigencias son diferentes en cada uno de estos casos.

En términos generales, los cálculos deben de realizarse aplicando un factor de seguridad de 1,5 para tensiones y corrientes.

Además, los cálculos se realizan atendiendo al riesgo de chispa en tres circuitos básicos: resistivo, capacitivo e inductivo.

Circuitos resistivos

Supóngase que se dispone de un circuito como el de la figura 31.6. ¿Qué condiciones debería satisfacer el circuito para que no hubiera riesgo de inflamación por chispa?

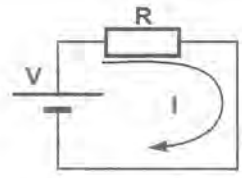


Figura 31.6. Circuito resistivo.

Claramente, el riesgo de chispa (por ejemplo, en caso de que se rompa un cable) dependerá de la tensión y la corriente en el circuito; de la energía manejada, en definitiva. Las normas [31.3] ofrecen la solución a la pregunta planteada incluyendo unas gráficas que indican la tensión y la corriente máxima que serían admisibles en función del grupo de gas (figura 31.7).

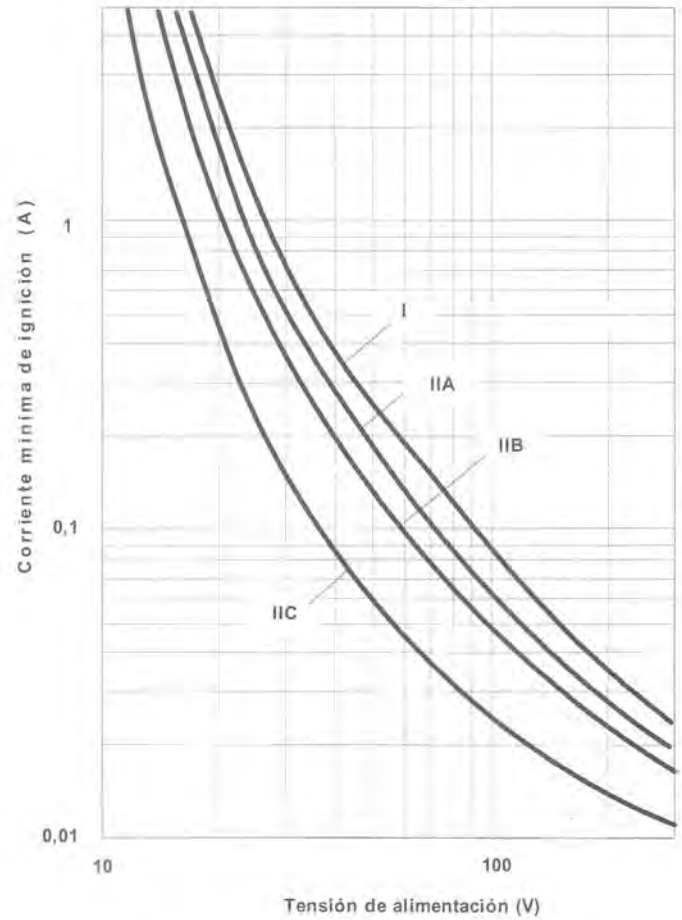


Figura 31.7. Tensión de alimentación y corriente mínima de ignición en un circuito resistivo según el grupo de gas (extraído de [31.3]).

EJERCICIO 31.3

Se dispone de un circuito resistivo como el de la figura 31.6 alimentado a una tensión máxima de 24 V. ¿Cuál sería la máxima corriente que podría circular por el circuito para que no exista riesgo de inflamación por chispa con un gas de grupo IIC? ¿Cuál sería la resistencia de menor valor óhmico y de tolerancia 5% que se podría utilizar?

SOLUCIÓN:

Según la figura 31.7, la máxima corriente permitida es de 261 mA para una tensión máxima de 24 V, lo que supone, aplicando un factor de seguridad de 1,5, una corriente máxima 174 mA (174·1,5 = 261 mA).

Por otra parte, para el cálculo de la resistencia hay que tener en cuenta que las normas obligan a tener en cuenta las tolerancias de los componentes. Suponiendo que la resistencia tiene una tolerancia del 5%, el peor caso tendría lugar cuando la resistencia tiene el menor valor posible, por lo tanto:

$$\frac{24\ \text{V}}{\text{R}(1-0{,}05)} = 174\ \text{mA}$$

Obteniéndose que la resistencia más pequeña es R = 145,2 Ω.

Un circuito electrónico está, normalmente, constituido de varias resistencias. Pudiera pensarse que la figura 31.7 provee un criterio para el cálculo de cada una ellas por separado. Sin embargo, es necesario considerar el conjunto como una carga completa cuya desconexión o conexión podría provocar una chispa. Por lo tanto, la figura es mucho más limitativa para el diseño de lo que pudiera pensarse. Así en principio, según el ejercicio 31.3, un circuito de seguridad intrínseca para el grupo IIC, no podría consumir más de 174 mA si la alimentación fuera como máximo de 24 V.

Circuitos capacitivos e inductivos

La limitación de la corriente en un circuito resistivo no garantiza la seguridad de un circuito debido a que la presencia de elementos que almacenan energía como condensadores y bobinas. Por ejemplo, supóngase que se dispone de un circuito constituido por un condensador como el de la figura 31.8, en el que se ha limitado la corriente de carga mediante una resistencia.

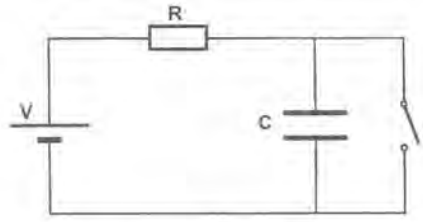


Figura 31.8. Circuito capacitivo. Un cortorcircuito del condensador podría provocar una corriente muy elevada por el interruptor aun cuando la corriente que proporciona la fuente de alimentación esté limitada.

En régimen permanente el condensador alcanzará la tensión V; un cortocircuito del condensador podría dar lugar a una corriente muy elevada y el consiguiente riesgo de chispa, dependiendo de la energía almacenada en el condensador ($\frac{1}{2}CV^2$), y por lo tanto, de su capacidad. Las normas indican, mediante unas gráficas, la tensión y la capacidad máxima en función del grupo de gas (figura 31.9).

De nuevo, la gráfica debe de aplicarse a la suma de todas las capacidades de un circuito. Hay que tener en cuenta que la mayor parte de los condensadores están situados entre la alimentación y masa, con lo que están en paralelo. Para un fallo consistente en el cortocircuito de uno de los condensadores, se estarían cortocircuitando realmente todos.

EJERCICIO 31.4

Se dispone de un circuito electrónico cuya tensión máxima de alimentación es de 15 V. La suma de las capacidades de todos los condensadores que componen el circuito es de 200 nF (incluyendo las tolerancias

para el peor caso, según indican las normas). ¿Cumple con los criterios de seguridad intrínseca para el grupo IIC?

SOLUCIÓN:

Aplicando un factor de seguridad de 1,5 para la tensión (15·1,5 = 22,5 V), se obtiene de la figura 31.9 que la máxima capacidad permitida para el grupo IIC es de 0,58 μF. Como la capacidad total es menor que este valor, se cumple con el criterio de seguridad.

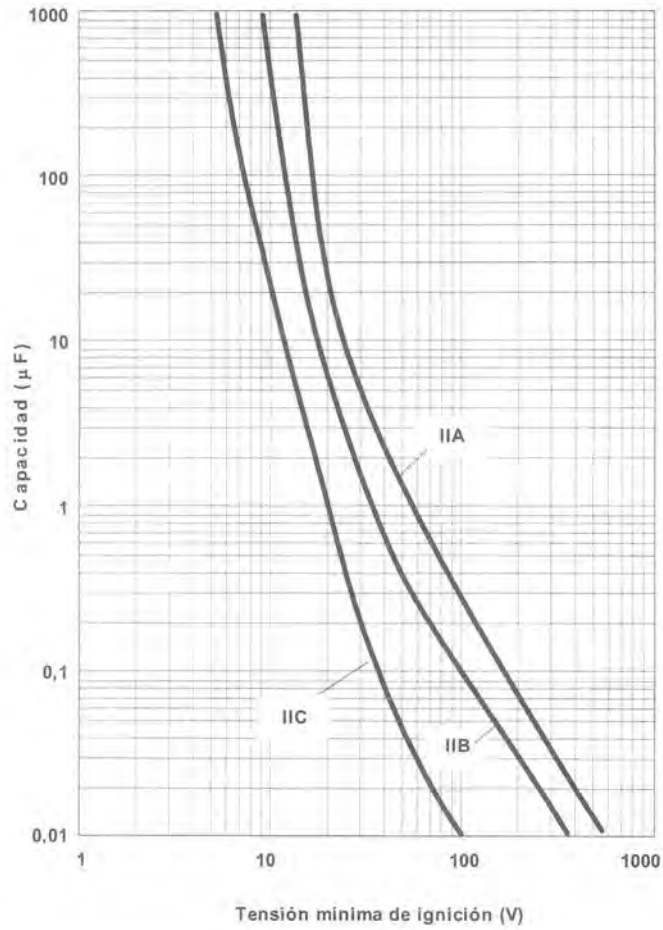


Figura 31.9. Capacidad máxima de un circuito en función de la tensión para los circuitos capacitivos del grupo II (extraído de [31.3]).

La limitación de la capacidad máxima puede ser excesivamente restrictiva para algunas aplicaciones. Pueden utilizarse capacidades mayores siempre que se limite de algún modo la máxima corriente de corto-circuito (por ejemplo, mediante una resistencia en serie). La norma EN50020 incluye otra serie de gráficas para los cálculos en estos casos.

Por otra parte, en los circuitos inductivos existen problemas similares debido al almacenamiento de energía ($\frac{1}{2}LI^2$). En estos casos, los peligros derivan de las sobretensiones que pueden producirse debido a la apertura

del circuito. Por ello, la norma también indica los valores de inductancia permitidos para cada grupo de gas en función de la corriente. En el caso de que el valor de inductancia necesario sea excesivamente elevado, los problemas pueden resolverse, normalmente, mediante la utilización de diodos en paralelo [31.6].

El material de seguridad intrínseca deberá marcarse, en general, con los valores máximos de tensión y de corriente de diseño, así como con los valores máximos de capacidad interna e inductancia interna.

31.2.2.2 Riesgo de inflamación debido a temperaturas elevadas

En el apartado anterior se han tratado, de forma general, los métodos de diseño para evitar que un equipo pueda inflamar un gas de un determinado grupo debido a la producción de chispas. Sin embargo, un gas también puede inflamar debido a que la temperatura de un componente sea muy elevada. Por ejemplo, en el ejercicio 31.3 se concluyó que la resistencia mínima del circuito que garantizaba que no existirían problemas relacionados con las chispas era de 145,2 Ω. Sin embargo, si esta resistencia tiene un tamaño muy pequeño podría calentarse mucho y alcanzar una temperatura suficiente como para inflamar el gas presente. Mientras que para evaluar los riesgos de chispas los gases se clasificaban en grupo I, IIA, IIB y IIC, para evaluar el riesgo por temperatura se clasifican en clase I, T1, T2..T6, como ya se ha mencionado. En principio deberá garantizarse que ninguna de las partes del equipo (desde los componentes hasta las pistas de circuito impreso) supere la máxima temperatura de la clase correspondiente.

Para facilitar el diseño, las normas indican cuál es la máxima corriente que puede circular por un cable de cobre de una determinada sección para que la temperatura sea inferior a la de inflamación dentro de la clasificación que corresponda. Por ejemplo, según las normas, para la clase T5, en un cable de cobre de diámetro 0,1 mm la corriente máxima admisible es de 1,9 A. Asimismo, se incluyen criterios similares para determinar las pistas de los circuitos impresos.

Por otra parte, el riesgo de inflamación no sólo depende de la temperatura, sino de la superficie del componente. Por ello, las normas permiten que pequeños componentes puedan superar la temperatura de inflamación con ciertas restricciones. Por ejemplo, para la clase de temperatura T5 (temperatura máxima 100ºC), se permite que un componente de menos de 10 cm^2 pueda alcanzar hasta 150ºC [31.6].

31.2.2.3 Interfaz entre circuitos de seguridad intrínseca y circuitos sin seguridad intrínseca

Los equipos de seguridad intrínseca situados en zonas peligrosas están generalmente unidos a equipos que están situados en zonas no peligrosas y que, por lo tanto, no presentan especiales requerimientos de seguridad. Por ejemplo, supóngase que se dispone de un sensor determinado (por ejemplo, de nivel) en un depósito que contiene un gas inflamable (figura 31.10) y que la visualización de los datos se va a efectuar en una sala

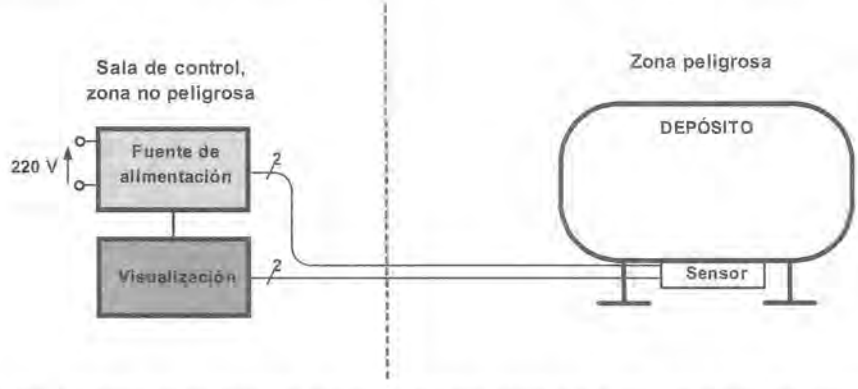


Figura 31.10. Un fallo en la zona no peligrosa podría conducir a una situación de explosión en la zona peligrosa aunque el sensor satisfaga todos los criterios de diseño de seguridad intrínseca.

de control, desde la que también se alimenta al sistema de medida mediante una fuente de alimentación continua. El sensor y la circuitería asociada pueden satisfacer todos los criterios de seguridad intrínseca pero, si debido a un fallo en la fuente de alimentación la tensión de la red eléctrica se aplicase directamente al sensor, se produciría una situación de claro peligro. Se necesita, por tanto, algún tipo de interfaz entre la zona segura y la peligrosa que limite la tensión y la corriente que puede llegar a esta última. Estas interfaces se conocen como sistemas de segregación.

Las interfaces más utilizadas son las barreras zéner. Con estos dispositivos se pretende limitar tanto la tensión como la corriente que puede llegar a la zona peligrosa. La estructura básica es la que se muestra en la figura 31.11.

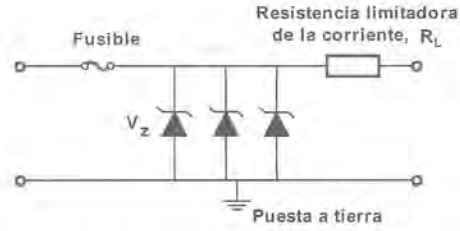


Figura 31.11. Esquema de una barrera zéner.

La barrera zéner está situada en la zona segura (figura 31.12). En caso de fallo, los diodos zéner y la resistencia limitan la tensión y la corriente, respectivamente. El fusible garantizaría la desconexión en caso de fallo. La razón de que la barrera de la figura incluya tres zéner es que, para satisfacer los criterios de seguridad intrínseca ia, se necesita que el dispositivo pueda ser seguro incluso con dos fallos. Por lo tanto, debe haber tres zéner para que, incluso aunque se estropeen dos, siga existiendo la limitación de tensión pretendida (para satisfacer los criterios ib bastarían dos zéner). De todas formas, cabe preguntarse por qué existe una única resistencia limitadora. El motivo es que la norma considera que determinados componentes son extraordinariamente seguros, de forma que si fallan lo hacen de tal modo que van a favor de la seguridad [31.3]. Estos componentes se denominan indefectibles y, entre ellos, se incluyen las resistencias de película y determinadas resistencias bobinadas debido a que si fallan quedan en circuito abierto con lo que se garantiza igualmente la limitación de corriente.

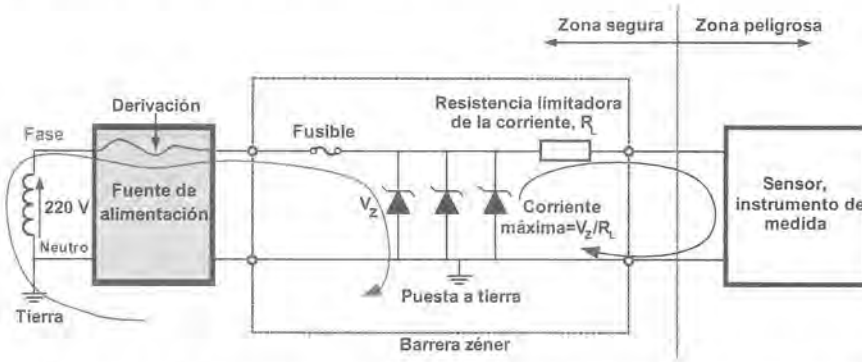


Figura 31.12. En caso de fallo, la tensión máxima en la zona peligrosa queda limitada por los zéner, y la corriente por la resistencia limitadora.

Las barreras zéner forman parte de lo que se denomina material eléctrico asociado que es aquel que contiene a la vez circuitos de seguridad intrínseca y sin seguridad intrinseca, y que se construyen de forma que los circuitos sin seguridad intrínseca no puedan afectar a los de seguridad intrínseca. El código de este material es similar al correspondiente al de seguridad intrínseca salvo que va entre corchetes; por ejemplo [EEx ia] IIC. Además, se indica cuál es la máxima capacidad y la máxima inductancia que se puede conectar a la barrera.

EJERCICIO 31.5

Para alimentar un circuito de seguridad intrínseca se dispone de una fuente de alimentación de 12 V. La barrera zéner está formada por tres zéner de 15 V. ¿Cuál debería de ser el valor de la resistencia limitadora para que la barrera pudiera ser denominada como [EEx ia] IIC?

SOLUCIÓN:

De las curvas de ignición de circuitos resistivos (figura 31.7) se obtiene que para 15 V, la máxima corriente permitida en el grupo IIC es de 1,35 A lo que supone, aplicando un factor de seguridad de 1,5, 0,9 A. Por tanto, la resistencia limitadora debería ser:

$$R_L > \frac{15\ V}{0,9\ A} = 16,67\ \Omega$$

Por otra parte, no se debe olvidar el efecto de la resistencia limitadora sobre la alimentación efectiva del circuito de seguridad intrínseca, puesto que provoca una importante caida de tensión que hay que tener en cuenta en el diseño.

31.2.2.4 La puesta a tierra

La normativa exige una puesta a tierra siguiendo rigurosos criterios. Además, en la zona peligrosa, los circuitos deben estar conectados a tierra por un solo punto. Supóngase que el circuito en la zona peligrosa está puesto a tierra por dos puntos: en el propio circuito y a través de la barrera zéner (figura 31.13). La Tierra, aunque sea un buen conductor, no es una superficie equipotencial. Por la Tierra circulan corrientes de derivación que dan lugar a que entre dos puntos diferentes exista, en la práctica, una caida de tensión impredecible. Por ello, en el circuito de la figura 31.13, por los conductores existentes entre la barrera zéner y el circuito de seguridad intrínseca pueden estar circulando corrientes de valores desconocidos y puede haber una tensión tal que no se cumplan los criterios de seguridad intrínseca al respecto. El problema se soluciona fácilmente si el circuito de seguridad intrinseca no se conecta a tierra (quedaría conectado únicamente a través de la barrera zéner). Por ello, las normas exigen que los circuitos puedan soportar un ensayo de rigidez dieléctrica respecto a tierra, de 500 V.

Sin embargo, en ocasiones resulta imposible no conectar a tierra el circuito de seguridad intrínseca. Por ejemplo, cuando con un termopar con la unión expuesta se desea medir la temperatura de un objeto haciendo contacto con él, se está realizando, en general, un contacto con tierra. El problema también existe con otros sensores, como los de conductividad de líquidos y las galgas extensométricas. Con estos circuitos resulta imposible cumplir con el ensayo de rigidez dieléctrica de 500 V. De todas formas el problema tiene varias posibles soluciones. Una consistiría en poner en contacto la tierra del circuito de seguridad intrínseca con la de la barrera zéner con un conductor especial de gran calibre de modo que se garantice que no hay caídas de tensión significativas entre las tierras. Otra solución, mucho más habitual en la actualidad, consiste en aislar

galvánicamente ambos circuitos (figura 31.14). De este modo, se pueden poner a tierra sin que la diferencia de tensiones entre las tierras pueda acarrear inconvenientes en cuanto que entre la barrera zéner y el circuito de seguridad intrínseca no existe contacto eléctrico (aparte de la propia tierra). Comercialmente existen diferentes modelos de interfaz con aislamiento que suelen incluir, además, circuitos de acondicionamiento para sensores, transmisores a 4-20 mA, etc. Por otra parte, puesto que los circuitos que no satisfacen el ensayo de rigidez dieléctrica deben instalarse de un modo especial, se marcan con una X en su código.

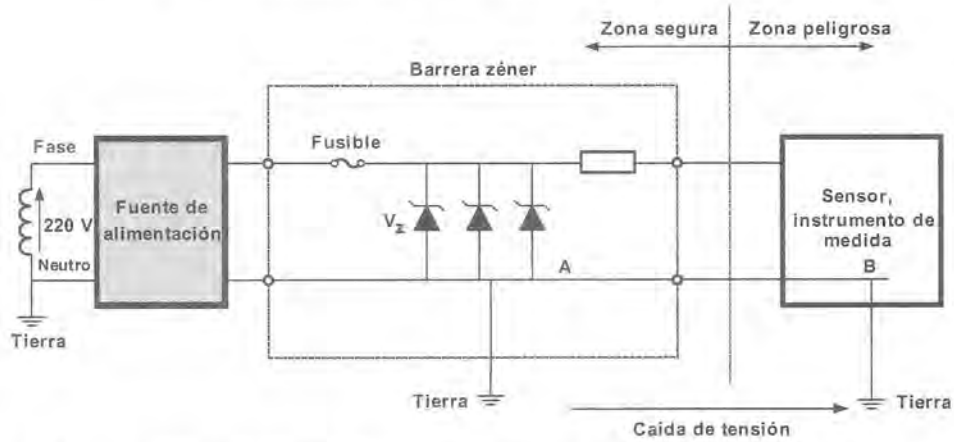


Figura 31.13. El instrumento situado en la zona peligrosa y la barrera zéner están puestos a tierra. Las diferencias de tensión entre ambas tierras podrían dar lugar a chispas si, por ejemplo, se rompiese el conductor AB.

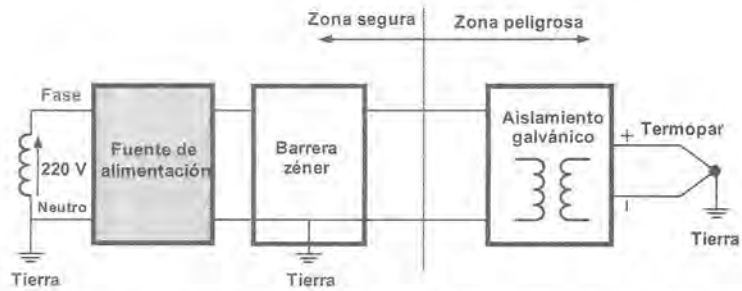


Figura 31.14. El aislamiento galvánico permite poner a tierra el circuito de seguridad intrínseca de forma independiente.

31.2.2.5 Certificación

Los equipos de seguridad intrínseca deben certificarse, generalmente, por medio de un organismo independiente que garantice el cumplimiento de las normas. Por ejemplo, en España, el Laboratorio Oficial José María de Madariaga (LOM) está acreditado para certificar la conformidad a las normas del material eléctrico para atmósferas explosivas.

Sin embargo, la norma considera que determinados componentes o conjuntos de componentes se pueden utilizar sin necesidad de certificado, aunque deben cumplir las exigencias de seguridad intrínseca. Dicho material se denomina simple y comprende:

- Componentes pasivos como interruptores, cajas de conexión, potenciómetros y dispositivos simples con semiconductores.
- Fuentes de almacenamiento de energía con parámetros bien definidos como condensadores o inductancias.
- Fuentes generadoras de energía (como termopares y células fotoeléctricas) que no produzcan más de 1,5 V, 100 mA y 25 mW.

Además, para que un material sea considerado como simple deben cumplirse determinados aspectos: por ejemplo, que la seguridad de un material simple no se obtenga por medido de otros dispositivos que limiten la tensión y/o la corriente. Así, una fuente de energía que produzca más de 1,5 V en la que se limite la tensión mediante un zéner para que no supere 1,5 V, no puede ser considerada como material simple (aun produciendo menos de 100 mA y 25 mW).

Entre el material no simple se encuentra el que contenga circuitos integrados, dispositivos piezoeléctricos, baterías, etc. [31.6] y deberá ser certificado para que pueda utilizarse en atmósferas explosivas. Además, cuando el material simple forma parte de un material que contenga otros circuitos eléctricos, el conjunto deberá ser certificado.

Resumen

En el diseño de cualquier sistema es necesario cumplir con los requisitos de seguridad que aparecen recogidos en las normas.

Los equipos electrónicos que se alimentan de la red eléctrica y tienen partes metálicas accesibles están, generalmente, conectados a tierra por motivos de seguridad para las personas.

Los equipos destinados a ser utilizados en atmósferas potencialmente explosivas deben cumplir normas muy rigurosas.

Un gas puede inflamarse porque se produzca una chispa o porque la temperatura de un objeto presente en la atmósfera sea suficientemente grande.

El sistema de protección de seguridad intrínseca consiste en que un equipo maneje una energía suficientemente baja como para no producir la inflamación de los gases presentes debido a una chispa y en que las temperaturas sean suficientemente bajas.

Los circuitos de seguridad intrínseca y sin seguridad intrínseca han de conectarse mediante algún tipo de interfaz que garantice la seguridad de la zona peligrosa. Una de las más utilizadas es la barrera zéner.

Los equipos de seguridad intrínseca deben certificarse, salvo el material simple.

Problemas propuestos

31.1 Un dispositivo de seguridad intrínseca viene marcado como EEx ia IIC T5. ¿Puede utilizarse cuando los gases presentes estén incluidos dentro de la clase de temperatura T6? ¿Puede utilizarse con un gas del grupo A? ¿Puede utilizarse en una zona 1?

31.2 Se pretende realizar un circuito electrónico para medir la concentración de gas metano en una mina de carbón mediante un sensor adecuado. Si la tensión de alimentación máxima que se considerará

para el diseño es de 28 V, ¿cuál debería ser el consumo máximo de corriente del circuito? ¿Cuál es el valor máximo de la suma de las capacidades de los condensadores que deseen emplear en el circuito sin recurrir a técnicas de limitación de corriente de descarga de los mismos? Si el circuito que se va a diseñar incluirá, entre otros, amplificadores de instrumentación, referencias estables de tensión, etc., ¿será necesario certificar el circuito?

31.3 Un dispositivo de seguridad intrínseca viene marcado como EEx ia IIC T6 (V_i = 30 V, I_i = 50 mA, C_i = 10 nF, L_i = 0). ¿Cuál es el significado de estos parámetros? ¿Puede conectarse una barrera zéner que viene marcada como [EEx ia] IIC (V_o = 35 V, I_o = 45 mA, C_o = 50 nF, L_o = 1 mH)?

31.4 ¿Necesita certificarse una caja de empalmes que vaya a ser utilizada en atmósferas explosivas?